RECONSTRUCTING PAST EVENTS:
A STUDY OF ENGINEERING FAILURE INVESTIGATIONS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHILOSOPHY
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Yafeng Wang
August 2020

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Helen Longino, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

David Hills

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Thomas Icard, III

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

George Smith

Approved for the Stanford University Committee on Graduate Studies.

Stacey F. Bent, Vice Provost for Graduate Education

Abstract

When a major engineering product failed, a failure investigation is often conducted to prevent similar failures in the future. In this dissertation, I propose an account of the epistemology and methodology of engineering failure investigations, based on a close examination of the documentations on five major plane crash investigations conducted by the National Transportation Safety Board (NTSB).

The dissertation is divided into three parts. The first part consists of the five case studies arranged in chronological order: the American Airlines Flight 191 accident in 1979; the United Flight 232 accident in 1989; the United Flight 585 accident in 1991; the USAir Flight 427 accident in 1994; and the TWA Flight 800 accident in 1996. In each case study, I summarize the entire investigation process, focusing on articulating the questions that arise and the evidential reasoning that helped resolve each question.

The second part of the dissertation examines how the investigators infer causes of failure events. The type of causal inference used in failure investigations typically proceeds from effects to causes, hence it is called reverse causal inference. This is in contrast with forward causal inference, where researchers start with an intervention and infer the effects of that intervention. I identify three types of reverse causal inference in engineering failure investigations: feature dependence, additional outcomes, and process tracing.

The third part of the dissertation examines how the investigators construct narratives of failure events. At the end of each failure investigation, the investigators come up with a narrative detailing the sequence of events leading to the outcome. This part of the dissertation examines how the investigators construct such narratives and support them with evidence. I argue that both the construction and the justification of a narrative of a failure event depend on the question-and-answer process in the investigation, which I call the question dynamics of the investigation. I examine three main components of the question dynamics: the resolution of questions, the significance of questions, and the arising of questions. I conclude the dissertation with an account of the coherence of narratives, which is a measure of the evidential status of narratives. My account of coherence relies on the question dynamics, and it captures the intuitive idea that the pieces of a coherent narrative “fit together” very well.

Acknowledgments

I want to begin by thanking the members of my dissertation committee. My dissertation advisor, Helen Longino, has been incredibly supportive and guided me throughout my entire dissertation writing process. Through numerous conversations, Helen helped me work through all the ideas in this dissertation. For instance, she encouraged me to pursue my idea of feature dependence long before I focused on studying engineering failure investigations. Helen has been a role model for me, both as a philosopher of science and as a teacher. Much of my style of doing philosophy of science reflects what I learned from her work and the philosophy of science courses she taught. I am grateful for having her as my advisor.

George Smith has been an extraordinary mentor to me ever since I studied at Tufts. Near a decade ago, his two-semester courses on Issac Newton's *Principia* had a profound impact on my intellectual growth and persuaded me to become a philosopher of science. More recently, George convinced me to focus on engineering failure investigations as my primary case studies. The plane crash cases he recommended me to study opened my eyes to the richness of the NTSB's investigative work. His decades of experience as a failure analyst helped me gain many insights about the practice. George is exceptionally generous with his time and encouragement, and I am fortunate to have him on my committee.

I also wish to thank David Hills for his extensive, careful, and insightful feedback on my ideas. David's comments, both oral and written, are invariably constructive. I first worked with David when writing my second-year paper on counterfactuals in history. His

critical feedback on my earlier drafts helped me significantly improve the quality of the paper. Similarly, his comments on an initial draft on reverse causal inference changed how I think about this topic. I appreciate everything I learned from him.

Finally, I had the opportunity to TA for Thomas Icard's logic courses, from which I learned a great deal. Thomas' breadth of knowledge and his ability to make intuitive ideas precise have always been inspiring. His "big-picture" questions about my dissertation helped me rethink the motivation behind the project. I also want to thank Thomas for his kindness and for reaching out to me during the COVID-19 pandemic.

The Stanford philosophy department is a wonderful community. I have greatly benefited from conversations with other students and visitors, including Rob Bassett, J.T. Chipman, Michael Cohen, Erin Cooper, Willie Costello, Huw Duffy, Blake Francis, Dave Gottlieb, Nathan Hauthaler, Hannah Kim, Dustin King, Olga Lenczewska, Christopher Lewis, Katy Meadows, Chris Mierzewski, Jared Palmer, Adwait Parker, Grace Patterson, Imran Thobani, Shane Steinert-Threlkeld, John Turman, Steve Woodworth, Francesca Zaffora Blando, and Adam Zweber. I also wish to express my appreciation of Jill Covington and the other administrative staff for their work.

I want to give a special shout out to the Leela Improv community for bringing so much fun and laughter to my life in the past few years. Thank you, Steven Burnett, Jill Eickmann, and Sidonie Starr, for being amazing improv teachers and coaches. Thank you, my improv friends and teammates Aaron Belkin, Alex Berg, Geraldine Carolan, Jason DeGraff, Eddie Dunoyer, Alyssa Frazee, Brett Glasner, Nancy Ma, and Galen Westerfield, for having my back on and off the stage.

Finally, I thank my dear friends Shuang Bing, Hongye Fan, and Abby Jaques for their support over the years. I thank my parents, Lianjun Yu and Xiaoyong Wang, for everything.

Contents

Chapter 1

Introduction

Event reconstruction research is a type of research that focuses on understanding how a past event or a series of past events occurred, based on the traces left behind by these past events. Researchers and investigators in a wide variety of fields engage in event reconstruction research: For instance, astronomical research on the formation of the Moon, geological and paleontological research on the extinction of the dinosaurs, historical reconstruction of the battle of Thermopylae, and engineering investigation of the inflight breakup of TWA 800 are all examples of event reconstruction research.

In recent years, philosophers of science have been increasingly interested in the epistemology and methodology of event reconstruction research. What makes event reconstruction research philosophically interesting is the fact that it is very different from paradigm cases of scientific research in theoretical or experimental sciences, and the fact that it faces unique epistemic challenges in reaching conclusions about the past.

First, unlike most of the theoretical and experimental researches that study *regularities*, an event reconstruction research studies an *individual* historical event with unusual or even unique characteristics. Moreover, part of the task of the event reconstruction research is precisely to reconstruct and to explain the unique aspects of the event in question. Consider the example of the origin of the Moon: The Earth's Moon is unique in the solar system because of how exceptionally large the Moon is; for instance, the ratio of moon-to-planet mass is around a hundred times larger for Earth than similar comparisons to the moons of Mars. One task of the astronomers researching the Moon's origin is to explain why the Earth's Moon is so huge compared to other moons in the solar system.

Second, unlike regularities that manifest themselves via reoccurring phenomena, an individual historical event is typically not repeatable in its entirety, and we often lack the practical means to recreate events that are similar to it in relevant aspects. For instance, part of the reason why Newtonian mechanics was so thoroughly tested is that it had been tested continually for over three hundred years, which was made possible by the stability of the celestial motions and the set of robust regularities underlying them. In contrast, the formation of the Moon only occurred once in the history of the solar system. So far, it is practically impossible to recreate events similar to it using physical experiments. Even for much smaller scale events such as plane crashes, investigators were limited in the aspects of the events they could recreate using flight tests and simulations.

Third, the event of interest often attracted the attention of the researchers only after it had happened; and in some cases, a long time after it had happened. Since the event itself occurred in the past and is no longer directly observable, our primary source of epistemic access to it consists of the event's causally downstream traces, which can be corrupted by factors outside of our control or simply decay over time. For instance, impact craters of asteroids could be displaced or destroyed by subduction of tectonic plates; fossils of dinosaur bones that were exposed by erosion would wear away over time; fracture surfaces from engineering accidents could be further damaged by contamination and corrosion; and historical records of human affairs might be biased or inaccurate.

Given the above limitations and challenges, how do researchers and investigators successfully reconstruct complex events of the past, and how do they support claims about past events with evidence? These are epistemological questions that many philosophers of science—myself included—are interested in. To address these questions, it is important to examine actual cases of empirical research that reconstruct past individual events. Most recent philosophical work on event reconstruction research has focused on case studies in geology, paleontology, and archaeology. In contrast, what is unique about this dissertation is that it draws methodological lessons about event reconstruction based on case studies of *engineering failure investigations.*

I choose to focus on engineering failure investigations as case studies of event reconstruction research for a few reasons. First, the epistemic challenges faced by engineering failure investigations tend to be less severe than many other types of event reconstruction research. Engineering failures tend to be simpler than most of the events studied in history and historical sciences such as geology and paleontology. Even the most complex engineering failures have a relatively short temporal duration and simpler causal structures than events such as the formation of the Moon, the Cretaceous-Paleogene (K-Pg) extinction event, or the Great Depression of the 1930s. Moreover, engineering failures tend to immediately precede the investigations, and there exist routine investigative protocols that help preserve as many traces from the engineering failures as possible. Consequently, challenges about the destruction and decay of traces are less severe in engineering failure investigations than many other types of research that reconstruct events of the deep past.

Second, the investigative processes of engineering failure investigations and the evidential reasoning underlying them tend to be better documented than in many other types of event reconstruction research. For instance, the National Transportation Safety Board (NTSB) is a major investigative agency responsible for investigating transportation accidents in the United States. It produces a very detailed accident report after it completes each investigation. The report provides details about all aspects of the accident, analysis of the factual data and causal conclusions, and safety recommendations based on the analysis. In addition, the NTSB typically holds public hearings after major milestones of each investigation, which are valuable sources of information about the questions that the investigators puzzled over and how they went about answering these questions. Finally, for high profile accidents, journalists who have insider access to the investigations also provide valuable insights into the investigative processes that are otherwise not easily visible to the outsiders.

Finally, engineering failure investigations—especially those conducted by the National Transportation Safety Board (NTSB) in the United States—have an excellent track record of *prima facie* successes of reconstructing past events. The *prima facie* successes of the NTSB accident investigations are evidenced by the following facts: First, the conclusions reached in the NTSB investigations are typically very *convincing* based on the evidential arguments advanced in the accident reports. Second, the conclusions of each NTSB investigation tend to enjoy a high degree of consensus among a large and very diverse group of experts within the investigation. Third, the safety recommendations based on the conclusions of the accident reconstruction tend to be very effective, in the sense that repeated failures are rare after the pertinent recommendations were properly implemented.

Therefore, even though this dissertation is motivated by epistemological questions about event reconstruction research in general, its case studies and philosophical discussions are entirely focused on engineering failure investigations. Ultimately, I hope that by gaining a better understanding of the methodological principles underlying engineering failure investigations, we can learn some useful lessons that are either applied to, or serve as contrasts with, event reconstruction researches in other fields such as geology, paleontology, archaeology and history. Nevertheless, my dissertation itself does not examine in depth the applicability of failure investigation methodology to other types of event reconstruction research, because this topic requires additional case studies that go beyond the scope of the dissertation.

This dissertation is divided into three main parts. The first main part (Part I) consists of five case studies of aviation accident investigations conducted by the NTSB, arranged in chronological order: the American Airlines Flight 191 accident in 1979; the United Flight 232 accident in 1989; the United Flight 585 accident in 1991; the USAir Flight 427 accident in 1994; and the TWA Flight 800 accident in 1996. In each case study, I summarize the entire investigation process, focusing on making explicit the questions that arise during the investigative process and the evidential reasoning that helps resolve each question. Reading through the case studies is not strictly required for understanding the remaining parts of the dissertation. However, it helps to get a sense of each investigation's full scale and the question-and-answer structure within it. Moreover, many details discussed in the case studies are also used to illustrate philosophical concepts and arguments in later parts of the dissertation.

The second main part (Part II) of the dissertation focuses on the methodology of *reverse causal inference*. When reconstructing a past event, the investigators often need to address questions about why certain sequences of events initiated, or why a given outcome was produced. Answering these why-questions requires making causal inferences to establish which causal factors initiated an event, or what events and conditions contributed to an outcome. Moreover, the type of causal inference commonly used in event reconstruction research has a "reverse direction", because it involves inferring from a given event or outcome to its causes. Hence I call this type of causal inference *reverse causal inference*. It is in contrast with the *forward causal inference* commonly used in experimental research, where researchers manipulate an independent variable and determine the effects of the intervention on a dependent variable.

How does reverse causal inference work? What are some common reverse causal inference patterns? I address these questions in Part II by introducing three types of reverse causal inference: Feature dependence, additional outcomes, and process tracing. The basic idea of *feature dependence* is that the features or details of an outcome are informative about the features of its causes; consequently, by choosing suitable features of an outcome, we can construct detailed profiles or descriptions of its causes. The basic idea of *additional outcomes* is that to determine whether a hypothetical event C causally contributed to a given outcome E, sometimes it is helpful to look for the *additional outcomes* of C to see if it indeed had occurred. Finally, the basic idea of *process tracing* is that to determine whether an event C causally contributed to another event E, sometimes it is helpful to trace through the causal processes leading from C to E. The bulk of Part II focuses on articulating the argumentative structures of these three types of reverse causal inference, using examples from the five case studies as illustrations.

Finally, the third main part (Part III) of the dissertation examines how the investigators come up with *narratives* of past events, and how they support these narratives with evidence. By a *narrative* of a complex past event, I mean an account of the causal etiology of this event. The narrative includes a chronology of the smaller "subevents" that constituted it, and an analysis of the causal relationships among these subevents and other background conditions. In other words, a narrative of a complex event captures not only the smaller events that constituted it, but also how these smaller events were connected, temporally and causally. The main task of Part III is to examine how event reconstruction research produces such a complex cognitive product and the main form of evidence in support of it.

My main thesis in Part III is that both the construction and the justification of a narrative of a past event depend on the question-and-answer process in the investigation, which I call the *question dynamics* of the investigation. The intuitive ideas behind this thesis are the following: First, the investigators in an event reconstruction do not come up with a narrative of a past event all at once. Rather, they answer one question at a time during the investigation, and continue to enrich a partial narrative based on the answers to previous questions. Moreover, the questions that arise in the investigation are related to each other in a structured way, partly because the questions that arise later tend to presuppose the answers to questions resolved earlier. Finally, how well the narratives of past events are supported by evidence depends on the extent to which the investigators can continue to satisfactorily resolve new questions that arise from the answers to previously resolved questions.

Part III attempts to make these intuitive ideas more precise by formally describing the abstract structure of question dynamics. A question dynamics has three main components: the resolution of a question, the significance of a question (relative to an investigation), and the emergence of new questions at each point during the investigation. Correspondingly, Part III contains three chapters addressing these three components, and the final chapter also combines all these components to represent the full structure of question dynamics. I conclude Part III with an account of the *coherence* of narratives, which is a measure of the evidential status of narratives, and it captures the intuitive idea that the pieces of a coherent narrative "fit together" very well.

Part I

The Five Case Studies

Chapter 2

Case Study 1: AA Flight 191

2.1 History of The Flight

At about 15:04 central daylight time (CDT), May 25, 1979, American Airlines Flight 191, a McDonnell-Douglas DC-10 airplane, crashed into a field less than a mile from its departure runway at Chicago O'Hare International Airport, Illinois. All 271 crew and passengers on board and 2 people on the ground died from the crash, making this crash the second deadliest accident in American commercial aviation history. ([8], p.2)

American Airlines Flight 191 was scheduled to depart from Chicago O'Hare International Airport to Los Angeles, California. On the day of the accident, AA Flight 191 departed from the gate at O'Hare International Airport. The temperature was 63 degrees Fahrenheit at the time. At 15:02:38 CDT, the control power cleared Flight 191 for takeoff on runway 32 right (32R). At 15:02:46, the captain acknowledged, "American one ninety-one underway." That was the last communication between Flight 191 and air traffic control. ([38], p.4)

During the aircraft's acceleration down the runway, the cockpit voice recorder (CVR) picked up the captain's voice calling out the plane's V-speeds. Everything sounded normal until the last stable takeoff thrust on the No. 1 engine was recorded two seconds before liftoff. One second later, the CVR recorded the word "damn", and it stopped recording.

At roughly the same time, witnesses in the control tower and on the ground saw that the left (No. 1) engine separated from the wing, went over the top of the wing, and fell on the runway. The rest of the airplane continued its takeoff. ([8], p.2) The air traffic controller tried to contact the captain, but there was no response. ([38], p.4)

Initially, the aircraft appeared to be climbing normally. After only 20 seconds in the air, however, it began a left turn. The turn got steeper and steeper, the nose of the airplane pitched down, and the plane dove towards the ground. During its descent, it continued to roll left until the wings were past the vertical position. ([38], p.4)

Flight 191 crashed in an open field near a trailer park about a mile from the departure end of the runway 32R. Because the plane crashed with a full load of fuel, an explosion and an enormous fire erupted upon impact. The crash killed everyone on board and two people on the ground. The pieces of the shattered plane damaged some trailer homes, and the smoke from the crash fire was visible from miles away. ([38], p.4)

2.2 Initial Investigation

2.2.1 Wreckage Recovery

The National Transportation Safety Board (NTSB) was among the first agencies to arrive at the crash site. After securing the crash site and its surrounding areas, the investigators spent about a week sifting through the wreckage and the pieces of the engine that fell off, placing tags on identifiable parts of the plane and surveying the scene for anything unusual. They recovered the two black boxes—the cockpit voice recorder (CVR) and the digital flight data recorder (DFDR)—and sent them back to the lab at NTSB headquarters to retrieve the data. Every day in the early investigation, the NTSB held a progress meeting among the investigators, and a separation informational meeting for the press and the victims' families. ([38], p.7)

The aircraft's wreckage was found at two locations: The left (No.1) engine, its pylon assembly[1] and a few other components were found at Runway 32R. The rest of the wreckage was located at the main crash site—the field and trailer park. The damage to the aircraft structure upon impact was so severe that the investigators found little useful information after an examination of the wreckage at the crash site. The No. 1 engine and pylon assembly that fell on the runway, in contrast, were largely intact. ([8], p.6)

The accident airplane, a McDonnell-Douglas DC-10, has three General Electric engines, one on each wing and one on the tail. The wing-mounted engines are coupled to pylons, a structure designed to carry the engines. The pylons are connected to the wings through four spherical joints. Two of the spherical joints align vertically in a forward bulkhead—an upright front wall attached to forward portions of the wing. A third spherical joint was at the upper surface of the pylon right behind the forward bulkhead. Its function is to transmit thrust loads from the pylon structure into a thrust link[2], which connects to the lower surface of the wing. Finally, there is a spherical joint in the aft bulkhead of the pylon, which attaches to a clevis[3] on the underside of the wing. ([8], p.6)

The investigators found the following structural elements either remaining with the separated No. 1 engine and its pylon or scattered along the runway: (1) The pylon's forward bulkhead; (2) the flange of the aft bulkhead—a projecting rim of the aft bulkhead that strengthens it and attaches it to the pylon; (3) the lower two-thirds of the aft bulkhead; (3) pieces of a bolt used to connect the thrust link to the pylon itself; (4) a 3-feet section of the left wing's leading edge near where the forward part of the pylon joined the wing. ([8], p.6-11) Since these pieces were the first structural elements of the airplane that failed, they had the potential to provide important clues to how the failure sequence initiated.

---

[1] A pylon serves to connect the frame of an aircraft to an object that it carries. Pylons for the wings are adaptors that attach the engine to the wings.

[2] The primary purpose of a thrust link assembly is to transmit the thrust generated by the engine to the airframe. The assembly is a sophisticated connecting rod designed with a special joint to reduce the transmission of twisting and bending loads.

[3] A clevis is a U-shaped piece that has holes at the end of its prongs.

2.2.2 Witness Interviews

Another source of information collected in the initial phase of the investigation was the witness reports. The witnesses agreed on the general shape of the story: The plane sped up down the runway, and, just about the time it took off, the left engine broke free and flipped over the wing. The rest of the plane continued its takeoff climb and shortly banked sharply to the left, eventually over 90 degrees. The nose dove and the plane crashed almost immediately. ([38], p.8)

Nevertheless, some witnesses provided more details than others. Several witnesses took pictures of the last seconds of Flight 191. The pictures provided data concerning the position of the control surfaces and the flight path. The investigators eventually used to corroborate the data recovered from the digital flight data recorder (DFDR). Many witnesses said that the separation of the engine happened during the plane's "rotation" —the few seconds of takeoff when the nose lifts the ground, but the rear wheels are still on the runway. This information helped to narrow down the timing of the initiation event. ([38], p.9)

Robert J. Graham, a maintenance supervisor for American Airlines, had arguably the best vantage point because he was waiting to cross runway 32R in a van when Flight 191 accelerated towards him. ([38], p.9) He stated:

> I noticed what appeared to be vapor or smoke of some type coming from the leading edge of the wing and the No. 1 engine pylon. I noticed that the No. 1 engine was bounding up and down quite a bit, and just about the time the aircraft got opposite my position and started rotation, the engine came off, went up over the top of the wing, and rolled back down on the runway. ([38] p.9)

The distribution of the wreckage and the witnesses' reports led to the conclusion that the failure sequence started with the separation of the No.1 engine-pylon assembly from the airplane's left wing. This preliminary conclusion gave rise to two important questions. First, why did the No. 1 engine-pylon assembly broke off? An engine falling off an airplane is an extraordinarily rare and unsettling event. As Robert Macintosh, a chief NTSB advisor, remarked, "Engines don't drop off by themselves." ([38], p.9) Second, why did the plane crash, given the loss of the No. 1 engine-pylon assembly? After all, DC-10s were designed to be flyable under catastrophic circumstances, even when one of the engines completely loses power. ([38], p.5) Why couldn't the pilot bring the airplane back down safety in this particular catastrophic circumstance? These two questions became the two main areas of the investigative focus.

2.3 Reconstructing the Engine-Pylon Separation

2.3.1 Fracture Analysis

Why did the No. 1 engine-pylon assembly break off? The engine showed no evidence of fire damage, so an explosion was ruled out early on. Witnesses agreed that the aircraft had not struck any foreign object on the runway. Also, no parts of the wing or other aircraft components were found along with the separated engine, other than the supporting pylon. These findings led investigators to conclude that nothing else had broken free from the airframe and struck the engine. Hence, the engine-pylon assembly separation could only have resulted from a structural failure ([38], p.9)

The next problem that the investigators needed to solve was determining the origin of the initial failure. Establishing the origin of structural failure is essential in failure analysis because the location of the origin determines which measures would prevent a repetition of the failure. Typically, the direction of crack growth (conversely, the direction towards the origin) can be detected from features of fracture surfaces such as chevron marks, crack branching, and river patterns. By examining these features, the investigators can trace the crack propagation back to the point of origin. Then they can investigate what initiated the crack at the origin—a corrosion pit, a porous region, or overloading, etc.

Investigators of the Flight 191 crash conducted a thorough examination of fractures at the pylon's attachment points to the wing: the forward bulkhead, the thrust link attachment, and the aft bulkhead. The deformation patterns and fracture characteristics at these points showed that the pylon's separation began with a fracture at the aft bulkhead, which allowed the aft end of the pylon to rotate downward due to the engine thrust. The thrust link attachment and the forward bulkhead then failed from overload, and the entire engine separated and flipped over the left wing. ([8], p.48)

The DC-10 aircraft pylon aft bulkhead connects to other parts of the pylon via a series of flanges—projecting rims around the periphery of the aft bulkhead. In the Flight 191 accident, the upper two-thirds of the pylon aft bulkhead separated from the flanges around its periphery and was found at the main wreckage site. In contrast, the upper flange, side flange, and the lower one-third of the aft bulkhead were found on the runway with the left engine-pylon assembly. When the investigators pieced together the fractured components, they discovered a 10-inch fracture on the upper flange with distinctive and informative features. ([8], p.12)

First, the investigators found chevron and tear marks on the fracture surface, which indicated that the rupture propagated downward at the center of the upper flange, then progressed both inboard (towards the fuselage) and outward (towards the wingtip). Moreover, the center portion of the fracture showed smearing characteristics of the compression portion of a bending fracture. In contrast, smearing became much less prevalent at the outer ends of the fracture. Together, these features suggest that the 10-inch-long fracture resulted from overload, which was initiated by a downward bending force applied at the center section of the upper flange. ([8], p.12)

Second, the upper flange's fracture surface also contained a crescent-shaped deformation, which exactly matched the shape of the bottom surface of the left-wing fitting clevis that couple with the bulkhead. Moreover, the investigators found a small, shallow dent in the fitting clevis's lower surface. The clevis was in a position that would line up horizontally with a fastener located at the upper flange surface, and the dent appeared to be caused by the fastener head hitting the clevis with a sliding movement. In short, there was strong evidence that the crescent-shaped deformation was caused by the wing clevis striking the upper flange in a downward direction. ([8], p.12-18)

Third, there were fatigue cracks at both ends of the fracture, and these fatigue cracks were informative about how the fractures continued to propagate after the initial 10-inch overstress fracture. For instance, at the inboard end of the overstress fracture, fatigue cracks continued to progress inboard and aft. When they reached the inboard side of the side flange, the cracks exhibited characteristics of rapid overstress; similarly at the outboard end. The overstress fracture and fatigue cracks were about 13 inches, and the rest of the fractures on the aft bulkhead (and on the pylon structures) all resulted from overstress. ([8], p.12) These findings suggest that it took some cycles of loads before the initial overstress fracture propagated enough to separate the pylon.

In short, features of the 10-inches fracture on the upper flange suggested that the fracture was initiated by the wing clevis striking and overstressing the upper flange in a downward direction. Fatigue cracks continued to grow from the initial overstress crack under cycles of takeoff and landing, eventually leading to the catastrophic separation of the engine-pylon assembly.

However, how did the wing clevis get into contact with the upper flange? When the investigators recovered the left wing clevis from the main wreckage site, it remained attached to the wing. Even though the aft bulkhead itself had fractured, upper portions of the aft bulkhead still attached to the clevis. To determine the clearance between the upper flange surface and the bottom surface of the wing clevis, the investigators reconstructed the relevant structure using the aft wing clevis fitting from the accident airplane and the aft bulkhead from another DC-10. They found that when the attachment pieces were in place connecting the clevis and the aft bulkhead, the vertical distance from the clevis's bottom to the upper flange's surface was about 0.5 inches. ([8], p.18) With the aft bulkhead to clevis attaching hardware in place, it would have been unlikely for the clevis to have been in contact with the upper flange. Conversely, in order for the clevis to have contacted the flange and created the crescent-shaped deformation, the attachment hardware connecting the clevis and the aft bulkhead would have been removed.

When the investigators removed the attachment pieces and moved the aft bulkhead upward, they found that the flange could be displaced about 0.6 inches above its previous position. In this position, the bottom of the clevis would be about 0.1 inches below the flange fracture's upper surface on the accident airplane. Moreover, the vertical depth of the crescent-shaped deformation on the accident airplane's aft bulkhead was precisely 0.1 inches. ([8], p.18) So the crescent-shaped deformation must have been created when the attachment pieces had been removed, and the aft bulkhead was pushed upwards. Since the attaching hardware was in place after the crash, the crescent-shaped deformation was not created during the accident. It must have been produced when the pylon was installed or removed from the wing during a maintenance operation. ([8], p.49)

2.3.2 Maintenance Records

The NTSB formed a maintenance group to examine the maintenance records of Flight 191. Investigators read through thousands of pages of documentation of Flight 191's 9-year maintenance history. They examined routine check-ups and filed complaints, reviewed DC-10's design and certification documents, McDonnell-Douglas' service bulletins, and American Airlines' maintenance procedures, looking for anything unusual or problematic. Soon, they made a discovery. ([38], p.10)

On March 29-30, 1979, about two months before the crash, the accident airplane had been at the American Airlines Maintenance facility in Tulsa, Oklahoma, for a routine check-up. During these two days, the aircraft underwent a special procedure to replace spherical bearings at the forward and aft bulkheads, a procedure that involved removing the engine and pylon from the wing. ([38], p.10)

McDonnell-Douglas had prescribed this operation a few years earlier in a service bulletin[4] sent out to all DC-10 operating airlines. In replacing the bearings, McDonnell-Douglas prescribed removing the engine from the pylon before removing the pylon from the wing. This way of dismantling the engine-pylon assembly required 79 disconnects of hydraulic and electric lines, and it could be time-consuming to carry out. So American Airlines and another major DC-10 operator (Continental Airlines) devised an alternative procedure that they believed to be more efficient than the that recommended by the service bulletin. ([2], p.1)

The new procedure adopted by American Airlines involved removing the engine and pylon as a unit. An engine stand and cradle were attached to the engine, and then a forklift supported the entire weight of the engine, pylon, engine stand, and cradle. After removing the pylon-to-wing attaching hardware, maintenance personnel lowered the entire assembly to access the spherical bearing. After the bearing replacement, they raised the entire unit by the forklift and reinstalled the attaching hardware. ([8], p.49) This alternative procedure had apparent advantages: For one thing, it saved almost 200 hours of maintenance per aircraft. More importantly, from a safety standpoint, it reduced the number of disconnects (of hydraulic and electric lines) from 79 to 27, decreasing the opportunities for damage or error. ([38], p.10)

When American Airlines contacted McDonnell-Douglas about the new procedure, the latter would not endorse it because reconnecting the combined engine-pylon assembly to the wing attachment points was too risky. The pylon without the engine weighs about 1,865 lbs, and the center of gravity is approximately 3 feet forward of the pylon-to-wing attachment points. In contrast, the pylon and engine assembly weigh 13,477 lbs, and the center of gravity is approximately 9 feet forward of the pylon-to-wing attachment hardware. When the engine is removed, the pylon is supported close to the pylon-to-wing attachment points,

---

[4] Technical service bulletins are recommended procedures for repairs and maintenance. They are used to officially alert airlines of problems with a plane and typically come with instructions on how to fix the problems. ([38], p.10)

and it is easier to observe and control motion between the pylon and wing structure. Thus, McDonnell-Douglas did not recommend removing the engine-pylon assembly as a unit. ([8], p.60)

The Federal Aviation Administration (FAA) regulations did not require airlines to get manufacturer approval for procedural changes. As a result, American Airlines continued with their procedure despite the manufacturer's refusal to endorse it. On July 28, 1978, about a year before the accident, American Airlines engineers issued an Engineering Change Order (ECO) about how to change the spherical bearings. It stated that a forklift with an attached engine stand must be positioned directly beneath the center of gravity of the engine-pylon assembly. That way, the pylon attaching points would support none of the unit's weight. ([2], p.1-2) The ECO also outlined a particular sequence of the tasks to be performed. According to this sequence, the forward attaching bearings are removed before the aft ones. However, the ECO did not caution that the tasks must be performed in the sequence's order. ([8], p.29)

Fast forward to March 29-30, 1979. When the accident airplane underwent the spherical bearing removal procedure at Tulsa, the maintenance work did not go smoothly. The midnight shift encountered great difficulty following the specific sequence outlined by ECO. So they disassembled the attaching hardware in the reverse order, removing the aft bearings before the forward bearings. ([2], p.2)

On the morning of 30th, the day shift arrived and noticed that the upper portion of the pylon's aft bulkhead was resting against the bolts attaching the clevis to the wing. Such a configuration required a 0.6-inch vertical upward movement of the aft bulkhead, which could only have occurred after the clevis deformed the upper flange. ([8], p.50) Instead of reporting this finding to their supervisors, the mechanics attempted to proceed with the replacement, but could not detach the forward bearings because the engine stand was not aligned. They shifted the forklift toward the front of the engine, removed the forward attaching hardware, and continued with the procedure. During this process, the forklift ran out of fuel and supported the unit underpowered, possibly causing a slight drift down of the lifting forks and their loads. ([2], p.2)

Based on these maintenance findings, the NTSB concluded that the overstress crack in the aft bulkhead's upper flange was introduced during the bearing-replacement operation on March 29 and 30, 1979. However, they could not determine the exact point in the procedure where the fracture initiated. As the NTSB report remarked, "a close examination of these maintenance procedures disclosed numerous possibilities for the upper flange of the aft bulkhead to be brought into contact with the wing-mounted clevis, and for a fracture-producing load to be applied during or after removal of the attaching hardware in the aft bulkhead's fitting." ([8], p.49)

One possibility involved operational errors such as misalignment of the forklift. Forklift operators were guided only by voice and hand signals, and they could not see the juncture between pylon and wing directly. The forklift positioning had to be precise; if the positioning were incorrect, the engine/pylon assembly would not be stable, causing it to rock back and forth and jamming the pylon against the wing's attachment points. Also, because of the close fit between the pylon to wing attachments and the minimum clearance between the structural pieces, even minor mistakes by the forklift operators while detaching or attaching the pylon could damage the aft bulkhead and its upper flange. ([8], p.49-50)

Another way the flange could be damaged was even more subtle: The forklift's forks could move slightly due to a pressure leak within the forklift's hydraulic system during the operation. The forklift used for Flight 191 ran out of fuel during the operation. When out of fuel, this model of forklift tended to drift downward by small amounts every hour. An unsuspected downward drift could have caused the flange to bear some weight of the engine-pylon assembly that exceeded its design limit, which could have been responsible for the overstress crack in the flange. ([38], p.13-14)

Even though the investigators could not determine which of these possibilities was responsible for the overstress crack on the upper flange, the presence of these possibilities supported the NTSB's conclusion that *some* steps in the bearing-replacement operation had caused the crack. To corroborate this conclusion, the investigators looked for cracked flanges in other DC-10s that had undergone the same operation. In addition, they zoomed in on another feature of the crack—its 10-inches length—and carried out various post-accident tests to determine the precise conditions that could produce overstress cracks of such a length.

2.3.3 Fleet Inspection and Post-Accident Testing

After the accident, the Federal Aviation Administration (FAA) conducted a fleet-wide inspection of the DC-10. During these post-accident inspections, 6 DC-10s had fractured upper flanges on the pylon aft bulkheads: four American Airlines DC-10s, and two Continental Airlines DC-10s. They went through the same maintenance procedure, during which engine and pylon were removed as a single unit. Metallurgists found that the failure modes on these planes were similar to that of the accident airplane. ([8], p.18) Moreover, the investigators learned from maintenance records that two other Continental Airlines DC-10s had had fractures on their upper flanges a few months ago. In these two cases, the damage was repaired per a method approved by McDonnell-Douglas. ([8], p.20) In short, the evidence from fleet inspection further corroborated that the upper flange overstress crack in the accident aircraft's aft bulkhead was created during a maintenance operation used by American and Continental Airlines. ([8], p.49)

Nevertheless, compared to the upper flange fractures found on these other DC-10s, the fracture found on the accident airplane had a distinctive feature: its length was 10 inches long, with an additional 3 inches of fatigue cracks extending from it. In contrast, the longest maintenance-induced crack found on the other DC-10's upper flanges was 6 inches. To understand how the 10-inch overstress crack was produced in the upper flange of the accident airplane, American Airlines and McDonnell-Douglas engineers carried out a variety of laboratory tests to reproduce it. ([8], p.50)

Since the upper flange crack might have been created by either static load (constant force, not in motion) or dynamic load (force in motion, such as striking), the investigators conducted both static loading and dynamic loading tests on flanges of the same type. The tests showed that a 6- to 7-inch crack was the longest that could be caused by deforming the flange with a single dynamic impact or static contact of the flange with the wing-mounted clevis. ([8], p.50) However, the tests conducted by McDonnell-Douglas showed that repeated load applications could produce a 10-inch crack in the upper flange. This result implies one possibility: The upper flange of the accident airplane contacted the clevis two or more times during the bearing-replacement operation. ([8], p.51) Given how tricky the procedure was, this possibility could not be ruled out.

Another explanation of the crack length proposed by American Airlines was that the crack occurred in two steps: First, a crack about 6 inches long was created during maintenance. Then, it grew to 10 inches upon the initial application of an abnormal operational load. American Airlines suggested that the aft bulkhead flange could have been subjected to an engine thrust load sufficiently great to extend the crack during takeoff. ([8], p.50) This suggestion was supported by a test conducted by American Airlines, in which a 6-inch crack was produced in the flange by forcing a wing clevis vertically down the pylon. When the flange was subsequently subjected to a thrust load, the 6-inch crack extended to 10 inches. ([8], p.21)

By design, the scenario suggested by American Airlines should not have happened. Given the designed clearance between the clevis and the bulkhead, nearly all the thrust loads were supposed to be transmitted to the wing through the thrust link. In contrast, the forward bulkhead and the aft bulkhead were only supposed to carry vertical and side loads. ([8], p.21) The investigators could not determine the pre-accident clearance within the aircraft pylon structure. However, they found other DC-10s during the fleet inspections, in which the clevis to bulkhead clearance was so small that a thrust load would have been imposed on the bulkhead and its flange. ([8], p.51) Hence the possibility suggested by

American Airlines could not be ruled out.

In sum, based on all the evidence from fracture analysis, maintenance records, fleet inspection, and post-accident testing, the NTSB concluded that the structural separation of the pylon resulted from a complete failure of the upper flange of the aft bulkhead. The upper flange's failure began with an overstress crack created five months before the accident during a maintenance operation. The crack extended to 10 inches, either by repeated errors in the maintenance or by service loads. In the months before the accident, the flange's residual strength continued to be reduced by the fatigue cracks caused by the takeoff and landing cycles of the aircraft. Finally, the flange was weakened so much that it failed when the accident aircraft applied full engine thrust for takeoff on May 25, 1979, causing a chain of overstress failures in the pylon to wing attachments that resulted in the engine's separation.

2.4 From the Engine-Pylon Separation to the Crash

2.4.1 The Black Boxes

Just two months after the crash of Flight 191, investigators had identified the major maintenance flaws and procedural deficiencies responsible for the separation of the engine. Even then, the investigation was far from complete. The DC-10 was designed to be flyable even if one of its engines lost power. To understand why the plane crashed after the physical loss of the No. 1 engine, the investigators needed to know what else went on in the plane before and after the engine separation, how the plane was flown, and the extent to which it was controllable.

One source of evidence that the investigators tapped into was the black box of Flight 191. There are two black boxes, and they are not black at all. Instead, they were painted fluorescent orange to make them easy to locate and recover from wreckage. One black box is the cockpit voice recorder (CVR), which records everything heard in the cockpit. The other is the digital flight recorder (DFDR), which, in 1979, recorded only a few flight parameters such as engine thrust, pitch angle, and altitude. ([38], p.19)

The DFDR also recorded the pilots' inputs for the airplane control surfaces[5] and how much the control surfaces moved as a result. The major control surfaces that are recorded include: the rudder, which is on the tail and controls the heading angle; the elevators, which is also on the tail and forces the nose to pitch up or down; the ailerons and flaps, which are on the wing surface and control the bank angle by stopping the airflow over a wing and forcing it down; and the slats, the wing extensions that give the wings more curvature and therefore more lift for takeoff and landing. ([38], p.16)

In the case of Flight 191, the investigators recovered both boxes. The DFDR recorded 50 seconds of data during the takeoff roll and 31 seconds of airborne data before the recording ended. The CVR recorded little data because it lost power and stopped recording when the engine separated from the airplane. The investigators could not hear what the pilots said during the next 31 seconds of the flight. Nevertheless, by correlating the two recordings, and with the help of other corroborating evidence, investigators were able to reconstruct a more detailed profile of the flight history. ([38], p.19)

At 15:02 CDT, May 25, 1979, Flight 191 received clearance for takeoff on runway 32R. The leading edge slats of the aircraft extended, giving the wings extra curvature to generate more lift. Lift—generating upward forces greater than the plane's weight—is a fundamental element of flight. When a plane speeds up under engine thrust, air flows over and under the plane's wings. Since the top of the wing has a greater curvature than the bottom, air travels faster over the top than the bottom, and the faster-traveling air has a lower pressure than the air underneath the wing. The result is an upward force on the wings. ([38], p.15)

As Flight 191 accelerated on the runway, the CVR recorded the pilots calling out a few pre-calculated takeoff speeds. Takeoff speeds are typically measured in "knots of indicated airspeed" (KIAS)—that is, how fast the plane moves relative to the surrounding air, because

---

[5] Control surfaces are aerodynamic devices on fixed-wing aircraft that allows them to move about three axes of rotation: (1) pitch, aircraft's nose up or down about an axis from side to side; (2) yaw, nose left or right about an axis from up to down; and (3) roll, rotation about an axis running from nose to tail.

the amount of lift a plane has depends on its speed relative to the air rather than the ground. ([38], p.15) The normal takeoff procedure identifies three critical speeds: V1, VR, and V2. V1 is the "decision speed" and represents the maximum speed at which the pilot can safely abort the takeoff. Above V1, there is not enough remaining runway to stop safely. After reaching V1 speed, it is safe to continue the takeoff even if one engine abruptly fails. VR, or rotation speed, is the speed at which the pilot pulls back the control column to raise the plane's nose. Finally, V2 is the "takeoff safety speed"; it is the speed that guarantees an acceptable rate of climb and other safety standards for aircraft controllability, even with one engine out. The three speeds are fundamental to safe flight and are pre-calculated as part of the flight plan. ([7], p.5)

The CVR on Flight 191 recorded the V and VR callouts by the captain after the DFDR recorded these speeds. The elevator deflected up at VR, and the aircraft rotated upward immediately. Flight 191 accelerated through V2 speed—153 knots in this case—during rotation[6]. Two seconds before the plane lifted off the runway, the DFDR recorded the last stable thrust reading for the No. 1 engine. One second later, the word "damn" was recorded on CRV, and then CVR ceased recording. ([8], p.4) simultaneously, the DFDR ceased recording the positions of the left inboard aileron, left inboard elevator, lower rudder, and two of the left-wing leading edge slats, all of which had been powered by the power generator associated with the now-separated No. 1 engine. ([8], p.5)

The plane lifted off at a speed of V2 + 6 knots relative to the air (KIAS), and it maintained a steady and normal climb for the first 9 seconds of the flight. During the climb, the DFDR reading for the No. 1 engine was zero; the No. 2 engine speed increased from 101 percent to a final value of 107 percent; and the No. 3 (tail-mounted) engine was at the takeoff setting. At the end of the climb, the aircraft accelerated to 172 KIAS and reached about 140 feet above ground level. ([8], p.5)

---

[6] Again, rotation here means the few seconds of takeoff when the nose lifts off the ground, but the rear wheels are still on the runway

However, the plane then slowed down. Eleven seconds later (and twenty seconds after liftoff), the plane had slowed down to 159 KIAS, and it began a left roll. The DFDR recorded right-rudder and right-wing-down aileron inputs from the pilots, which means that the pilots tried to use these control surfaces to force the right-wing down. However, the left roll increased despite the increasing right rudder and right-wing down aileron deflections. Three seconds before impact, the plane was banking 90 degrees to the left and perpendicular to the ground. By the time it crashed, it had rolled over onto its back and was diving nose-first to the ground at 21 degrees, with full counter aileron and rudder controls and nearly full up elevator being applied. ([8], p.5)

2.4.2 Hydraulic System Damage and Retraction of the Left Outboard Slats

Why did Flight 191 begin a left roll 20 seconds into the climb and continue to roll until the crash, despite the pilot's efforts to level the wings? The DFDR records contained a few clues: A decrease in airspeed preceded the left turn; the turn was not a controlled move since the pilots' inputs suggested that they were doing their best to counter it; the aircraft descended, and its nose dropped before it crashed into the ground. These are classical signs of a stall, which happens when an aerofoil cannot make enough lift to keep the aircraft in flight. What went wrong with Flight 191 seemed to be that its left wing stalled (whereas its right wing did not).

The investigators soon found out why the left wing, but not the right wing, stalled. A 3-feet section of the left wing's leading edge skin was found on the runway with the pylon structure; it was torn away by the forward bulkhead of the pylon when the latter separated from the left wing. Moreover, several hydraulic lines routed through the left-wing leading edge were severed because of the engine-pylon separation, which caused a portion of the left-wing leading edge slats to lose hydraulic pressure and to retract. ([8], p.11)

The leading edge slats are extensions on the leading edge of the wings that can give the wing more curvature and, therefore, more lift. During a normal takeoff, the leading edge slats of the airplane wings must be locked in an extended position to generate proper lift; otherwise, the wings will stall. Like other modern aircraft, the leading edge slats and other control surfaces on the DC-10 are controlled by the hydraulic systems. The extension and retraction lines of the leading edge slats connect to hydraulic cylinders located behind them. When the slats are extended and the control valve is closed, hydraulic fluid is trapped in the operating lines and the hydraulic cylinders. The incompressibility of this fluid holds the slats extended against any external air loads. ([8], p.54) Unlike other aircraft designs, however, the DC-10 did not include a separate mechanism to lock the extended leading-edge slats in place, relying solely on the system's hydraulic pressure. ([2], p.1)

A DC-10 has three hydraulic systems. If one hydraulic system fails, the other two are sufficient to power the control surfaces themselves. To provide the redundancy to cope with a hydraulic system failure, most of the control surfaces are powered by two hydraulic systems. The left-wing leading edge slats, for instance, are connected with both the No. 1 and the No. 3 hydraulic systems. Unfortunately, when the pylon separated from Flight 191's left wing, both the No. 1 and No. 3 hydraulic systems' operating lines for the outboard portion of the left-wing leading edge slats were severed. ([8], p.11) As the hydraulic fluid trapped in the operating lines was lost, the outboard portion of the leading edge slats on the left wing slowly retracted, since there was less and less pressure holding them extended and the force of the oncoming air pressure was pushing them back in. ([38], p.16)

The leading edge slats' uncommanded retraction had a dangerous effect on Flight 191. Post-accident tests showed that with the slats extended, the speed at which the wings stalled was about 124 knots, whereas the stall speed increased to 159 knots when the slats were retracted. ([8], p.54) With the left outboard slats retracted, the lift of the left wing was reduced, and the left wing stalled when the airspeed of the plane decreased to 159 knots. In the meantime, the rest of the plane still had its hydraulic pressure. The slats on the right wing were extended, and so the right wing did not stall. Consequently, the plane entered an uncommanded left turn. Examination of the wreckage of the leading edge slats confirmed that upon impact, the left wing's outboard slats were retracted. In contrast, the left wing's inboard slats and the right wing's inboard and outboard slats were extended to the takeoff position. ([8], p.11)

2.4.3 The Speed Reduction and the Pilots' Response to Stall

The retraction of the left wing outboard slats was not the only contributing factor to the left wing's stall. The DFDR data showed that Flight 191 was climbing normally for the first 9 seconds of the flight, with an airspeed higher than its left wing's stall speed. The pilots then decreased the power of the No. 2 engine, and the left wing only stalled when the plane's airspeed reduced to 159 knots. This prompts a question: Why did the pilots reduce the engine power and decrease the airspeed in the first place?

Furthermore, entering a stall should not by itself cause a plane to crash. As an illustration, paper airplanes stall and recover on their own: the familiar pattern of a paper plane swooping up to a stop then entering a dive is nothing but a series of stalls and recoveries. When pilots react to the onset of a stall promptly, commercial jets can recover from stalls in similar ways. The procedure for stall recovery is familiar to any commercial pilot: immediately point the nose down and increase the engine power. If the pilots do this at the onset of the stall, the plane will gain enough speed to fly again. ([38], p.17)

However, the DFDR data showed that the pilots did not pitch the nose down or increase the engine power after the onset of the stall. Instead, they continued to pull back the control yoke and tried to pitch the plane's nose up, even as the plane descended because of the stall. Before the DFDR stopped recording, it recorded that the elevator had increased to the full nose-up deflection. ([8], p.5) The pilots' effort was in vain: By pulling back on the yoke and pitching the plane nose up, they only slowed down the plane and further decreased the lift. Without gaining enough speed to regain lift, it was hopeless for Flight 191 to recover from the stall. ([38], p.17) This begs a second question: Why didn't the pilots follow the standard stall recovery procedure after the left wing's stall?

To answer these two questions, The NTSB investigators examined the emergency procedures in American Airlines' Operating Manual for the DC-10, particularly those relating to an engine failure during takeoff. They also examined the effects that the engine's separation would have on the aircraft's electrical and instrumentation systems for understanding which instruments were working in the cockpit and what information was available to the pilots. ([8], p.54)

The answer to the first question—why the pilots reduced engine power in the first place—became clear upon examining American Airlines' emergency procedure: The pilots were following the carrier's prescribed engine failure procedure during takeoffs. The procedure assumed that the pilots received warnings of an engine failure after the plane had already reached V1 speed, in which case the pilots were required to continue the takeoff. Moreover, the procedure called for the plane to climb out at V2 speed until reaching 800 feet or the obstacle clearance altitude, whichever is higher. ([8], p.45)

The emergency engine failure procedure fits well with the accident airplane's flight profile. First, The CVR of Flight 191 recorded V- and VR-callouts with no incident; and it lost power one second before the plane lifted off the runway, at which point the plane was already accelerating through V2 speed. By the time the pilots realized that the No. 1 engine had failed, they would have to continue the takeoff. ([8], p.4)

Second, the pilots could not see the No. 1 engine and left wing from the cockpit. ([8], p.24) Instead of witnessing the engine separation, they would have seen the power meter for the No. 1 engine dropping to zero, and an "engine out" warning light indicating the loss of power in the No. 1 engine.[7] So the pilots would probably have assumed the No. 1 engine lost power due to a failure, not that the No. 1 engine separated from the wing physically. ([38], p.17) Therefore, it would have been natural for the pilots to follow American Airlines' engine failure procedure, which was designed more for engine power loss than for engine

---

[7] The investigators established that these instruments still worked after the engine separation.

separation.

Third, when Flight 191 lifted off the ground, its speed was 159 knots relative to the air (KIAS), and it was still accelerating. By the end of the first nine seconds, the aircraft accelerated to 172 KIAS and reached an altitude of 140 feet above ground level. ([8], p.5) Since the emergence procedure called for the aircraft to fly at its V2 speed—which was 153 KIAS—until 800 feet was reached, it would make sense for the pilots to comply with the procedure and reduce the power of the remaining engines.

American Airlines' takeoff engine failure procedure had a rationale behind it. When an engine loses power, a series of events follows. First, the other engine or engines go into overdrive. Normally, for takeoff, the engines' power settings are about 80 percent of their full capability. However, when one engine loses power, the others compensate and reach levels above 100 percent. For instance, during the first 9 seconds when Flight 191 was accelerating, the DFDR reading of the No. 2 engine speed increased from 101 percent to 107 percent. ([8], p.5) The idea behind the procedure was that if the plane were to keep flying at high speed, the other engines might fail because of overwork; hence the pilots should slow the plane down to the takeoff safety speed (V2) during the plane's climb. ([38], p.18)

Despite the good intentions behind the emergency procedure, however, the effects of following it in Flight 191 were disastrous. During its deceleration toward V2, the aircraft reached 159 KIAS—the stall speed for the left wing—and the uncommanded left roll began. The pilots did not follow the standard stall recovery procedure and, in the end, could not get the plane out of the stall. This brings us back to the second question mentioned earlier: Why didn't the pilots follow the standard stall recovery procedure after the onset of the stall?

One possible answer relates to the altitude of Flight 191. When the left roll began, Flight 191 was 20 seconds into the flight and 325 feet above ground level. ([8], p.5) At such a low altitude, if the pilots had tried to recover from the stall by entering a dive to gain speed, the result might have been an immediate impact with the ground. The low altitude also gave little time for the pilots to react and take corrective actions. ([38], p.18)

The NTSB investigators, however, suspected that there was more to the story than the low altitude. They knew that the CVR lost power, and the DFDR stopped recording certain flight parameters when the No. 1 engine separated from the left wing. This suggested that the engine's separation had caused damages to the aircraft's electrical systems, which could have made a difference to which instruments were working in the cockpit and what information was available to the pilots during the 31 seconds of the flight. The power loss of the CVR before takeoff meant that the investigators could not hear what the pilots had said during the flight. At the same time, it provided clues about the electrical situation in the cockpit: If the CVR lost power, then everything else on that circuit must have also lost power. That was the starting point to figuring out exactly what was and was not working in the cockpit during the flight. ([38], p.19)

There are three electrical systems on the DC-10, each powered by one of the three engines. Each engine provides mechanical energy to its associated generator, which transforms the energy into electricity. The generator's electrical power is distributed to individual electrical components of the aircraft. The generators can function either in parallel or in isolation, and each generator can supply enough power to operate all essential electrical components on the aircraft. Moreover, the generator output can also charge the aircraft batteries. Batteries are used for aircraft startup, and as an emergency source of power in an electrical system failure. One backup power source on Flight 191, for instance, can provide about 30 minutes of emergency power for the captain's instruments, essential communication instruments, and navigation equipment. ([8], p.43)

In addition to three separately powered systems for redundancy, there are layers of protective circuitry that automatically isolate problems, so they do not spread and damage the rest of the system. For an analogy, the isolation mechanism works just like a circuit breaker in a house. When one circuit gets overloaded, the circuit breaker cuts off the current supply to that part of the house, ensuring that the current does not overheat the wires in the rest of the house. ([38], p.19)

Similarly, protective circuitry on the DC-10 automatically isolates faulty electrical components from other parts of the system. When there is a current surge in a generator, the protective circuitry will engage in a "lockout mechanism" that isolates the faulty generator and all of its wirings from the rest of the airplane's electrical system. Any components powered by the generator lose power as a result, but the rest of the electrical system continues to function. ([38], p.19)

When the No. 1 engine fell off Flight 191, the generator that the engine had powered also failed. Furthermore, the separation of engine and pylon also severed electrical wire bundles near the pylon, which included the main circuits between the engine generator and its associated electrical components. Since the No. 1 engine generator had powered the CVR and portions of DFDR, the loss of power in these devices provided evidence that the No. 1 engine generator's lockout mechanism came into effect, probably as a result of short circuits during the engine separation. ([8], p.52) However, the No. 1 engine generator had powered many other aircraft systems and instruments, including the left stall warning computer, the stick shaker[8] on the captain's control yoke, and the slat disagreement warning system. As a result of the lockout mechanism, these instruments would also have lost power. ([8], p.44)

The additional loss of these flight instruments had disastrous consequences for Flight 191. First, the slat disagreement warning light should have illuminated after the uncommanded retraction of the left wing leading edge slats. However, because it was inoperative, the flight crew would not have received a visual warning of the slat asymmetry. Second, when the left wing approaches stall speed, the left stall warning computer activates both an auditory alert and the stick shaker motor on the captain's control yoke. However, since the left stall warning computer and the stick shaker were powered by the generator driven

---

[8] The stick shaker causes the yoke to shake when the plane is entering a stall, so the pilot, whose hands are on the yoke, cannot help but notice the stall warning. ([38], p.20)

by the No. 1 engine, both devices became inoperative after losing that engine. ([8], p.67) Moreover, the first officer's control yoke was not equipped with a stick shaker (which, if installed, would have been powered by a different engine generator). McDonnell-Douglas offered to add a stick shaker for the first officer, but American Airlines decided not to install it on its DC-10 fleet.

Losing these warning systems created a situation that gave the flight crew few opportunities to recognize and prevent the aircraft's ensuing stall. To make the situation worse, the aircraft configuration itself also provided little or no sign of the stall onset. The left outboard slats retracted while inboard slats remained extended. Therefore, the stall would be limited to the outboard segment of the left wing, and the flight crew would felt little or no buffet. The DFDR also indicated some turbulence, which could have masked any aerodynamic buffeting from the stall. Finally, the roll began at 159 KIAS, which was well above the aircraft's normal stall speed (124 KIAS). So the pilots probably did not suspect that the roll to the left indicated a stall. The roll likely confused them, especially since the stick shaker did not activate, and the only information available to them suggested that they were dealing with an engine failure and a little turbulence. ([8], p.54; [38], p.18)

In short, the pilots did not follow the standard stall recovery procedure after the onset of the stall, because the situation they were in made it exceedingly difficult for them to recognize the stall. Most significantly, the lockout mechanism on the No. 1 engine generator caused the critical stall warning systems on the aircraft to lose power, depriving the pilots of the primary source of stall signals.

The NTSB investigators discovered that the lockout mechanism on Flight 191 could have been unlocked. Suppose the flight crew had unlocked the mechanism. In that case, they could have restored power to the stall warning instruments with power generated from the other two engines. However, the phrase "could have" in the preceding sentences needs to be qualified. Besides the captain and the first officer, there was a flight engineer in the cockpit. The lockout mechanism could only be released if the flight engineer had flipped a switch on his electrical and generator reset panel. The problem was that when the flight engineer was positioned for takeoff, he could not reach this panel. Instead, he must re-position his seat to face the panel and get out of his seat to reach the switches. ([8], p.44) The NTSB concluded that the flight crew probably did not restore the lost electrical power, either because they were too preoccupied with regaining control of the plane, or because there was not enough time (11 seconds from stall to crash) for them to evaluate and respond to the electrical emergency. ([8], p.52)

2.4.4 Wind Tunnel Tests and Flight Simulations

Was there anything else that the pilots could have done to avoid the crash? To what extent was Flight 191 controllable after the engine-pylon separation? To better answer these questions, the investigators conducted a series of wind tunnel and flight simulator tests. First, researchers at the National Aeronautics and Space Administration (NASA) used the wind tunnel to determine the aerodynamic characteristics of a DC-10 with the left engine and pylon missing and the left wing's outboard leading edge slats retracted. Those tests confirmed that the outboard slats' retraction increased the stall speed of the left wing from 124 knots (143 mph) to 159 knots (183 mph) relative to the air. ([8], p.23)

Next, the DFDR data, aerodynamic data from wind tunnel tests, and the weather conditions during the accident were incorporated into a Motion Simulator at McDonnell-Douglas. The tests simulated the separation of the No. 1 engine-pylon assembly and its aerodynamic effects, the retraction of the left wing's outboard leading edge slats, and the loss of the No. 1 and No. 3 hydraulic systems. Moreover, some of the tests simulated the loss of the stall warning system and its stick shaker function, whereas other tests simulated the retention of these systems. Thirteen pilots took part in the flight simulation tests, and they were all thoroughly briefed on the flight profile of Flight 191. ([8], p.23)

The flight simulation tests yielded the following major results: (1) In all cases, the roll began at 159 KIAS, confirming that 159 KIAS was the stall speed of the left wing after the outboard slat retraction. (2) In all cases where the pilots applied the control inputs recorded in Flight 191's DFDR, the aircraft lost control, and the simulations ended in nearly identical crashes. ([8], p.23) (3) Whenever the airspeed was above 159 KIAS, the DC-10 was controllable regardless of the engine loss, the slat retraction, and the damages to the No. 1 and No. 3 hydraulic systems. (4) On those tests runs where the pilots immediately followed the standard stall recovery procedure upon the start of the left roll, they could recover from the stall and continue the flight. ([8], p.24)

Finally, The FAA conducted a second series of flight simulation tests to determine the takeoff and landing characteristics of the DC-10 with an asymmetrical leading-edge slat configuration. In these tests, the slat disagreement light and stall warning systems, which provided warning based on the 159 KIAS stall speed, were programmed to operate normally. None of the pilots experienced any problems with the aircraft controllability for either takeoff or landing during these tests. The FAA simulator tests further supported the conclusions that the pilots of Flight 191 could have flown the aircraft successfully at speeds above 159 KIAS. If they had recognized the roll onset as a stall, they could have lowered the nose, and accelerated the aircraft out of the stall. ([8], p.25)

However, all the successful simulator flights shared a common characteristic: The pilots all recognized the left roll as the onset of the stall, because the stall warning system was functioning, and the test pilots knew about the circumstances of Flight 191 before. The test pilots agreed that based upon the accident circumstances and the lack of stall warning systems on the accident airplane, it was not reasonable to expect the pilots of Flight 191 to recognize the beginning of the roll as a stall or to recover from the roll. The NTSB concurred. ([8], p.54)

In summary, The NTSB determined that the loss of control of Flight 191 was caused by the combination of three events: the retraction of the left wing's outboard leading edge slats; the loss of the slat disagreement warning system; and the loss of the stall warning system. The three events all resulted from the separation of the engine-pylon assembly. The simulation tests showed that each event by itself would not have caused a qualified flight crew to lose control of its aircraft. However, together during a critical portion of the flight, they created a situation that made it extremely challenging for the flight crew to recognize and prevent the aircraft's ensuing stall. ([8], p.55)

2.5 Conclusion

The NTSB concluded that Flight 191 crashed because it entered an asymmetrical stall and left roll. The stall was caused by an uncommanded retraction of the left wing outboard leading edge slats. The pilots did not recognize the onset of the stall because of the loss of the slat disagreement indication and stall warning systems. The failures of these systems resulted from the separation of the No. 1 engine and pylon assembly from the left wing at a critical point during takeoff, which resulted from damage to the pylon structure by improper maintenance procedure. The NTSB identified a variety of contributing factors to the accident, including deficiencies and vulnerabilities in the design, maintenance, and operational procedures of the DC-10s. It made safety recommendations to address these issues accordingly.

First, the investigation revealed several design flaws of the DC-10s. One design flaw is the lack of redundancy of the slat locking mechanism. Although a mechanical locking device for the leading edge slats was standard on other aircraft, McDonnell-Douglas did not include this feature on the DC-10, opting to use pressure within the hydraulic system as the sole locking mechanism for the leading edge slats. ([2], p.3) After this accident, the commercial aviation industry universally adopted mechanical slat locking devices. With this new design, once the slats are extended, they will stay that way until the pilot voluntarily brings them back. ([38], p.22)

Another design flaw is the lack of redundancy of the stall warning systems. The DC-10 had one stall warning system for each wing. When the stall warning computer detected a stall in its respective wing, it would cause a stick shaker motor to vibrate the captain's control yoke. The problem is that the left and right systems lacked crossover information, and the right system could not detect a stall in the left wing and vice versa. Furthermore, only the captain's control yoke has a stick shaker motor, and loss of power to that motor prevented it from shaking the control yoke. ([2], p.3) Therefore, the NTSB called for a design change that allowed crossover between the two stall warning systems, so that the stick shaker could be activated either by the left or the right computer. Furthermore, the NTSB recommended installing two separately powered stick shaker motors on each DC-10, one for the captain's control yoke and the other for the first officer's control yoke. ([2], p.4)

Second, the NTSB also identified multiple deficiencies in the maintenance procedures and practices, as well as deficiencies in the communications among the operators, the manufacturer, and the FAA regarding maintenance damage incidents. For instance, using a forklift to remove the engine-pylon assembly is highly problematic. The tolerances of pylon attaching hardware are tiny, and it is extremely challenging to manipulate a forklift bearing 13, 477 lb load with extreme accuracy. ([2], p.3) Hence, the NTSB argued that airlines should immediately discontinue the practice of removing the engine and pylon as a single unit. ([8], p.70)

The investigation also highlighted vague standards in airline reporting requirements regarding maintenance damages. For instance, during a pylon removal procedural for the accident airplane two months before the accident, the maintenance crew found the pylon bulkhead resting against the wing clevis in an abnormal configuration. However, they were not required to report this finding to their supervisors. For another example, Continental Airlines had performed an identical procedure to remove pylons on its DC-10s. They discovered two cases of maintenance-induced damage on the pylon bulkheads months before the Flight 191 accident. However, because of the latitude in FAA's guidelines for reporting major repairs, Continental Airlines classified the bulkhead damage as "minor" despite its repeat occurrence, and chose not the report the incident to the FAA. The manufacturer

McDonnell-Douglas, when receiving notification of nearly identical bulkhead damage, simply accepted the "minor" maintenance damage classification. The lack of rigorous reporting criteria resulted in multiple lost opportunities to discover procedural flaws. ([2], p.3)

In response, The NTSB recommended the FAA to more clearly define "major" and "minor" repair categories so that any repair to damage to a structurally significant component would require reporting. In addition, the NTSB argued that appropriate FAA personnel should be required to evaluate the damage reports to determine whether the damage indicates unsafe practice, so that proper actions are taken to disseminate relevant safety information to other airlines and maintenance facilities. ([8], p.72)

Finally, the NTSB concluded that American Airlines' engine-out emergency takeoff procedure required re-evaluation. The pilot's adherence to the takeoff-climb airspeed prescribed in the company's engine-out emergency procedure resulted in the aircraft entering the stall speed phase of flight. Had the pilot maintained excess airspeed during its climb, the accident might not have occurred. Since the airspeed schedules in American Airlines' emergency procedures at the time of the accident were identical to those contained in the emergency procedures of other air carriers, the Safety Board recommended re-calculation of the speed schedules for the engine-out climb to ensure that they provide the maximum protection from the stall. ([8], p.54)

Chapter 3

Case Study 2: United Flight 232

3.1 History of the Flight

On July 19, 1989, at about 16:00 central daylight time (CDT), United Airlines Flight 232, a McDonnell-Douglas DC-10-10, crashed landed near Sioux City, Iowa. Even though 111 of the 296 passengers and crew died from the crash, the flight crew's performance in the accident was remarkable, given a large number of survivors and how the flight crew handled the emergency without any hydraulic flight controls. ([6], p.168)

United Airlines Flight 232 was scheduled to depart from Denver, Colorado, to Philadelphia, Pennsylvania, with a stop at Chicago, Illinois. The flight crew consisted of Captain Alfred Haynes, first officer William Records, and flight engineer Dudley Dvorak. Dennis Fitch, an off-duty United Airlines DC-10 captain and flight instructor, was on board as a first-class passenger. ([4], p.1) United Flight 232 departed Denver at about 14:09 CDT and reached its cruising altitude of 37,000 feet without issue. ([6], p.169)

About 15:16:10 (about 1 hour and 7 minutes after takeoff), the flight crew heard a loud bang, and the plane vibrated and shuddered. From the engine instrument panel, the flight crew noticed that the No.2 (tail-mounted) engine had failed, and the flight crew went through United Airlines' engine shutdown checklist for the No.2 engine. ([11], p.1)

Meanwhile, the airplane entered a descending turn to the right, and the first officer (the pilot flying) announced that the airplane did not respond to his control column inputs. The first officer turned the control column all the way left and pulled it all the way back, which commanded maximum left aileron[1] and maximum up elevator[2] inputs. However, the airplane continued to pitch down and bank to the right. The captain tried to control the airplane with his control column, but the airplane again failed to respond to the inputs. At this time, the right bank increased to 38 degrees. To prevent the airplane from rolling into an inverted position, the flight crew closed the No. 1 (left) engine throttle and firewalled (i.e., applied full power of) the No. 3 (right) engine, and the airplane slowly returned to the wing-level attitude. ([25])

Furthermore, the flight engineer reported that all the hydraulic pressure and quantity gauges had dropped to zero, which meant that all the three independent hydraulic systems on the airplane had failed. ([11], p.1) Captain Haynes described the consequences of such a triple hydraulic failure as follows:

> That left us at 37,000 feet with no ailerons to control roll, no rudders to coordinate a turn, no elevators to control pitch, no leading edge devices to help us slow down for landing, no trailing edge flaps to be used in landing, no spoilers on the wings to slow us down in flight or to help to brake on the ground, no nosewheel steering and no brakes. That did not leave us a great deal to work with. ([24])

The simultaneous failure of all three hydraulic systems on a DC-10 had been regarded as virtually impossible, and nothing in the training of the flight crew had prepared them for it. When the flight engineer radioed the San Francisco United Airlines Maintenance and reported the loss of all hydraulic systems, he was told that there were no established procedures for such an event. ([22], p.23)

Besides the total lack of hydraulic control, the flight crew had two other problems.

---

[1] Ailerons are flight control surfaces on each wing's outer rear edge. They control the aircraft's roll.

[2] Elevators are flight control surfaces on the aircraft's tail, which control the aircraft's pitch.

First, the airplane began to oscillate longitudinally (i.e., pitch up and down) following the initial dive and roll. The oscillations (called "phugoids") took between 40 to 60 seconds per cycle. In each cycle, the plane would pitch down while accelerating, and then pitch up while decelerating. In a normal flight, such oscillations in pitch and airspeed naturally damp out after small adjustments ("trim") of the flight control surfaces (the elevator in particular). In the case of United flight 232, however, the only thing the flight crew could do was adjusting the power settings of two remaining engines. They could never eliminate the oscillations or regain pitch control. ([24])

Second, the airplane tried constantly to roll to the right. Three times, the plane rolled up to 38 degrees to the right and was close to flipping over, and the flight crew had to apply asymmetric thrust on the two engines to level the wings. ([6], p.170) To maintain a constant heading, the flight crew kept adjusting the engine power, and the airplane flew at high speed. ([4], p.2)

The flight crew was learning on the fly about how to use only engine thrust adjustments to control all the essential aspects of the flight, such as pitch, roll, airspeed, and vertical acceleration. ([22], p.23) Many of the required new flying skills were counter-intuitive: For instance, the pilots discovered that when the nose of the airplane pitched up, they must reduce the throttles to dampen the pitching motion, which was the opposite of what they had been trained to do to avoid a stall. ([6], p.170) Moreover, even though the control column inputs did not seem to accomplish anything, letting go of the control column was very psychologically challenging for the pilots. So the captain and the first officer had their hands on the control columns, and only released the control columns when they needed to make several quick adjustments to the No.1 and No.3 engine throttles to help control the flight. In the meantime, the flight engineer was on the radio, trying to get more help. ([24])

At about 15:20 (about 4 minutes into the accident), the flight crew radioed the Minneapolis Air Route Traffic Control Center (ARTCC) and declared an emergency. Initially, ARTCC suggested Des Moines International Airport for an emergency landing. However, the flight crew decided that the airport was too far away (over 170 miles), and they could not keep the airplane in the air for that long. The ARTCC controller then suggested Sioux Gateway Airport (SUX) at Sioux City, Iowa, the nearest suitable airport (about 70 miles away). The flight crew confirmed the destination and were given vectors to the airport. ([11], p.1-3)

About 15:29 (about 13 minutes into the accident), a flight attendant informed the captain that Dennis Fitch, an off-duty DC-10 flight instructor on board, had volunteered to offer help. The flight instructor was immediately invited into the cockpit. After a few minutes of conversing with the flight crew, the flight instructor decided that he did not have new suggestions. The captain asked the flight instructor to observe the ailerons through the passenger cabin windows to see if control column inputs had any effects on the ailerons. The flight instructor reported back that the ailerons were not moving at all. Nevertheless, the captain and the first officer continued to manipulate their control columns, hoping for at least some effect; and they asked the flight instructor to take over control of the engine throttles. ([11], p.3) So the flight instructor took one throttle in each hand and made rough steering adjustments based on what the other pilots were doing with the control columns. ([6], p.171)

About 15:34, the flight crew decided to attempt a no-flap, no-slat landing[3] at Sioux Gateway Airport. The flight engineer contacted the air traffic controller at the Sioux Gateway Airport for information about the runway for an emergency landing. The controller suggested runway 31, which was 9,000 feet long. At that point, the airplane was about 35 miles northeast of the airport. ([11], p.22)

To prepare for the emergency landing, the flight crew dumped fuel to reduce the airplane's weight, extended the landing gears via an alternative mechanical procedure, and advised the flight attendants to prepare the passengers for an emergency landing. ([11],

---

[3] In a normal landing, the slats on the leading edge and the flaps on the trailing edge of the wings are extended, which increased the lift on the wings and allowed the aircraft to operate at lower speeds. Without hydraulic pressure to deploy flaps and slats, the aircraft had to maintain a high level of speed during landing.

p.3) Because the airplane tended to roll to the right, it was much easier to turn the airplane in that direction, and so the flight crew executed a series of right turns to reduce altitude and line up with the airport. ([6], p.171)

About 15:58 (about 2 minutes before impact), United flight 232 came out of its last right turn. However, because of the engine throttles' limited control, the flight crew could not line up with runway 31. Instead, the flight lined up with runway 22, which was closed and only 6,600 feet long, but ended in an open field. Given the difficulty in maneuvering, the captain decided to continue the approach to runway 22 instead of runway 31. ([11], p.23) Because the air traffic control had expected a landing on runway 31 and had placed fire trucks on runway 22, the controller had to move the vehicles out of the way before the airplane touched down. ([25])

On the final approach to landing, the flight crew continued to control the aircraft by adjusting the engine thrust. To land safely, the plane must fly straight and level, and its forward and downward velocities must be within safety limits. Normally, a safe DC-10 landing would have a forward velocity of 140 knots (160 mph) and a downward velocity (sink rate) of about 3-5 feet per second. Because of the loss of all hydraulic systems on United flight 232, however, the slats and flaps on the wings that would have helped the airplane slow down for landing could not be deployed. As a result, the accident airplane attempted to land at almost 220 knots (250 mph), with a sink rate of over 30 feet per second. ([6], 171-172)

About 100 feet above the ground, flight 232 began one of its down-phugoids. The nose of the airplane pitched down, the rate of descent increased, and the right wing dropped about the same time. The flight crew tried to correct this by adding engine power, but they did not have enough time to react. About 16:00 (approximately 44 minutes after the failure of the No.2 engine), the accident airplane hit the ground on runway 22. ([11], p.5)

The first impact came from the right wing tip followed by the right main landing gear, which gouged an 18-inch-deep hole in the concrete runway. The No.3 (right) engine burst into flames after contact with the ground, and the main fuselage caught fire because of the spilled fuel from the right engine. The cockpit, the tail, and the right wing all broke off as the plane skid along the runway, and the fuselage broke into three sections. The center section (rows 9-30), which contained most passengers, rolled into an inverted position and finally came to a stop in a cornfield near the end of the runway, about 3,700 feet from the initial impact. ([6], p.172)

3.2 Emergency Response

Back in 1987, Sioux Gateway Airport conducted a full-scale disaster drill to simulate the crash of a wide-body aircraft with 150 passengers on the closed runway 22. The drill found some shortcomings in the airport's emergency response plan, particularly the lack of personnel and equipment due to the small size of the airport. In response, the airport improved its emergency plan by incorporating people and resources from the neighboring communities. ([25])

On July 19, 1989, the Sioux City Airport received a warning at 15:25 (about 35 minutes before the crash) that United flight 232 was heading towards Sioux City. The airport quickly mobilized all the emergency services at its disposal. All the local hospitals, clinics, and health centers were notified, and a plane took off in Des Moines, heading to Sioux Gateway Airport with more medical supplies. Moreover, July 19 was the day of the month when the Air National Guard at Sioux City was on duty, and 285 trained National Guardsmen participated in the emergency rescue mission. The emergency vehicles, firefighters, and medical personnel arrived at their designated positions promptly, preparing for an emergency landing at runway 31. When it became clear that United flight 232 could not make it to runway 31 and would crash land on runway 22 instead, all the emergency vehicles and personnel relocated under 2 minutes. ([25])

After the airplane crashed landed and the main pieces of its fuselage came to rest, a fire burned along the center section that contained most passengers, and the firefighters launched a barrage of extinguishing foam as surviving passengers emerged from the burning wreckage. ([22], p.25) However, the height and density of the cornstalks, wind direction, and the failure of a water pump during water resupply attempts limited the firefighter's ability to control the post-crash fire. Soon afterward, the fire spread within the cabin and could not be reached by an exterior firefighting attack, and the fire burned out of control for over two and a half hours. ([11], p.94)

Most passengers managed to walk through the ruptures in the fuselage structure and away from the burning cabin. Each survivor was assigned a group of medical care personnel and brought to the local hospital by ambulance. ([25]) As for the crew, for over 35 minutes, they were trapped in a waist-high crumpled cockpit because the rescue personnel who saw the cockpit remnant assumed that anyone in it was dead. After the flight crew regained consciousness, the flight engineer found a hole in the wreckage's aluminum skin and waved out of it to attract attention. ([22], p.25) Eventually, the rescue personnel used a fork-lift to pry apart the cockpit, pulled the flight crew out of it, and brought the four injured crew members to the local hospital. ([25])

Of all the 296 people on board, 185 (including all the four crew members) survived the crash and the post-crash fire; 76 died from blunt-force injuries sustained in the ground impact; 35 died from smoke inhalation in the fire. ([11], p.35)

3.3 Initial Investigation

The National Transportation Safety Board (NTSB) investigated the United flight 232 crash. It zoomed in on the most remarkable fact about the accident—that all the three hydraulic systems on the aircraft had failed.

Before the accident, A triple hydraulic failure was thought to be nearly impossible. McDonnell Douglas, the FAA, and United Airlines considered the failure of all hydraulic systems so unlikely that they developed no procedure for this situation. When the United flight 232 crew called United Airlines' maintenance headquarters for assistance, they had to repeat that they lost all hydraulics because the experts had difficulty believing it. ([6], p.177)

Nevertheless, it was indisputable that a total hydraulic failure had occurred in the United flight 232 accident. Shortly after the No.2 (tail-mounted) engine failure, the flight crew noticed that the hydraulic fluid pressure and quantity gauges had dropped to zero in all the three systems. About 1 minute after the engine failure, the Flight Data Recorder (FDR) recorded no further movement of the flight control surfaces. The flight crew communications recorded by the Cockpit Voice Recorder (CVR) further confirmed that the control surfaces did not respond to the cockpit inputs. ([11], p.75)

Why did all the hydraulic systems on United flight 232 fail? From the start, the NTSB's search for the triple hydraulic failure's causes centered on the No.2 engine. The total loss of hydraulics happened right after the failure of the No.2 engine, and there was evidence that the engine failure inflicted substantial damage on the tail of the accident airplane, through which the hydraulic lines of all the three systems were routed. During the flight, the flight engineer went to the passenger cabin to inspect the tail visually, and he reported that both the right and left horizontal stabilizers[4] were damaged. ([11], p.3) Moreover, Sioux City residents had photographed the accident airplane as it approached the airport, and the photographs showed large holes on the right horizontal stabilizer and missing components of the No.2 engine[5]. ([11], p.7) Could it be that some rotating parts of the No.2 engine burst in flight and the high-speed fragments damaged the hydraulic lines routed through the horizontal stabilizers?

To evaluate this possibility, the NTSB investigators conducted a three-dimensional reconstruction of the accident airplane's tail in a hanger at Sioux Gateway Airport. ([11],

---

[4] Horizontal stabilizers are horizontal surfaces on the tail that provide stability for the pitching motion of the airplane.

[5] Specifically, the fan cowl door and the tail cone of the No.2 engine were missing. ([11], p.7)

p.30) During the reconstruction, the investigators found severed pressure lines of the No.1 and No.3 hydraulic systems in the right horizontal stabilizer. Moreover, X-ray energy dispersion examination of the material adhering to the fracture surfaces showed traces of titanium alloy. ([11], p.35) The only components near the right horizontal stabilizer made from titanium alloy were the No.2 engine components, including the fan blades and fan disks. Therefore, the investigators concluded that fragments released during the No.2 engine failure severed the No.1 and No.3 hydraulic system lines, and drained all the hydraulic fluid in the two systems' reservoir. ([11], p.75)

The No.2 hydraulic system failed differently. Unlike the No.1 and No.3 hydraulic systems, the No.2 hydraulic system was powered by the No.2 engine. The engine-driven hydraulic pumps of the No.2 hydraulic system were below the fan section of the No.2 engine. Shortly after the accident, farmers in a rural area near Alta Iowa found the hydraulic supply and return hoses from the No.2 engine-driven hydraulic pumps, together with fan blade fragments and a few other engine components adjacent to the hydraulic pumps. ([11], p.24) Since the investigators could not find any other system anomalies that would explain the failure of the No.2 hydraulic system, they concluded that the No.2 hydraulic system failed because of the separation of the hydraulic supply and return hoses from the aircraft, which in turn was caused by the fragments released from the No.2 engine. ([11], p.75)

In short, all three hydraulic systems failed because of high-speed fragments released from some rotating parts of the No.2 engine. The No.2 engine of the accident airplane is a General Electric CF6-6 model, and it consists of four major sections: The fan section (front), the compressor section, the combustion section, and the turbine section (aft). ([11], p.13) All the four sections had rotating components. Which rotating components of the No.2 engine had burst in flight and initiated the failure sequence?

The evidence pointed to the fan section in front of the engine. The fan section of the No.2 engine consists of a stage 1 fan disk and attached fan blades, a stage 2 fan disk and attached fan blades, the spinner cone and cover, and various mounting and balancing hardware.

([11], p.41) Most conspicuously, except a few stage 1 fan blade fragments embedded in the horizontal stabilizers or on the ground at the airport, the entire stage 1 fan disk was missing from the crash site. Moreover, the stage 1 fan disk was made from titanium alloy, the material found on the fracture surfaces of the severed hydraulic lines. For these reasons, the stage 1 fan disk almost immediately became a prime suspect. The investigators believed that the fracture of the No.2 engine stage 1 fan disk was the initial event of the failure sequence. ([11], p.75)

To find out *how* and *why* the No.2 engine stage 1 fan disk had fractured, the investigators needed to recover the fractured fan disk. Based on the FDR, CVR, and radar data, the investigators estimated when the fan disk fractured and the aircraft's state of motion at that time. After calculating the likely trajectory of the fan disk debris to find out where it could have landed, they narrowed the possible locations down to a cornfield near Alta, Iowa, about 36 kilometers in size. The investigators and local volunteers scoured the area but found that the corn stocks concealed nearly everything on the ground, and they could not locate the pieces after weeks of search. General Electric, the manufacturer of the fan disk, offered large sums in rewards for critical parts of the disk. ([32])

About three months after the accident, an Iowa farmer found a section of about two-thirds of the stage 1 fan disk when harvesting corn. A day later, another farmer found a smaller section of the fan disk about a mile away. ([32]) The two sections made up nearly the entire disk, each with fan blade segments attached. The locations of two sections of the stage 1 fan disk relative to the radar track of flight 232 confirmed that they were the first parts of the engine to separate from the aircraft. Moreover, the smaller segment of the stage 1 fan disk departed the aircraft to the left, and the rest of the fan disk assembly departed to the right. ([11], p.25) In short, the investigators were correct in assuming that the fracture of the stage 1 fan disk was the initial event of the failure sequence. The main question then became: Why did the stage 1 fan disk of the No.2 engine break apart in the first place?

3.4 Fracture Analysis

The stage 1 fan disk is a machined titanium alloy forging that weighs about 370 pounds. It is about 32 inches in diameter, and its cross-sectional shape is divided into four portions: the rim, the bore, the web, and the disk arm. The rim is about 5 inches thick and contains slots that retain the fan blades. The bore is about 3 inches thick and is next to the 11-inch-diameter center hole of the disk. The web is about 0.75 inches thick and extends between the rim and the bore. Finally, the disk arm extends aft from the web and is bolted to a rotating shaft. ([11], p.41)

The fan disk from the accident airplane contained two main fractures, which caused one-third of the rim to separate from the disk. One fracture extends mostly circumferentially[6] through the web and the rim, and the other fracture extends mostly radially[7] through the bore, the web, and the rim. To understand how these two fractures came about, the NTSB investigators conducted detailed examinations of the physical characteristics, chemical compositions, and microstructures of the fracture surfaces. ([11], p.41)

Analysis of the features on the circumferential fracture showed that they were typical of overstress[8]. The near radial, bore-to-rim fracture also contained overstress features over most of its surfaces; however, the overstress features stemmed from a pre-existing fatigue crack[9] region in the bore of the fan disk. ([11], p.45) The investigators inferred that the fan disk fracture and separation began with the fatigue crack in the bore area. They reconstructed how the fatigue crack had led to the fan disk separation using stress analysis: After growing to a sufficient size, the fatigue crack created a bending moment that overstressed

---

[6] i.e., perpendicular to a radius of the circular disk.

[7] i.e., parallel to a radius of the disk.

[8] Overstress fractures occur during one force application and at loads equal or greater than the yield strength (i.e., stress needed to cause the material to deform) of the material. The load that caused the overstress fracture was applied in a millisecond before the component failed. ([33], p.57)

[9] Fatigue fractures typically occur at loads less than the yield strength of the material. Usually, they require many cycles of stress to initiate and to grow before the component fails. For the fan disk, a stress cycle is one takeoff and landing where the disk spins up to its design speed and is fully loaded by centrifugal force. ([6], p.190)

the disk, causing two large overstress fractures to form almost instantaneously. The overstress fractures allowed a disk segment about one-third of the disk to exit the engine. As soon as the disk segment was released, the rest of the disk was immediately out of balance and exited the engine in the opposite direction. ([11], p.77)

The next step of the investigation was determining the spatial and temporal origin of the fatigue crack. The investigators found that the fatigue crack initiated near a small cavity on the disk bore's surface. The cavity is 0.055 inches wide (about the size of the period at the end of this sentence), and 0.015 inches deep into the bore surface. ([6], p.198) Moreover, metallurgical and chemical analysis of the region surrounding the cavity revealed a defect, which was an altered titanium microstructure known as the hard alpha. ([11], p.45)

All molten metals form crystal structures on a microscopic level when cooled below the melting point. Titanium has two crystal structures (or two "phases"[10]), alpha and beta. The alpha phase tends to be harder and more brittle, and the beta phase tends to be more ductile. In pure titanium, the alpha phase is stable below 882 degrees Celsius, and it transforms into the beta phase above 882 degrees Celsius. By adding certain alloying elements to pure titanium, the transformation temperature could be lowered to the point where the beta phase could be present even at room temperature. The titanium alloy used in the stage 1 fan disk, Ti-6Al-4V, contains nearly equal quantities of alpha and beta phases. As a result, Ti-6Al-4V strikes a good balance between strength and ductility and maintains its properties at high temperatures. ([22], p.27)

The hard alpha defect surrounding the cavity in the fan disk bore, however, consisted of the pure alpha phase. Its brittleness was further elevated by the presence of an excessive amount of nitrogen. Since titanium can only absorb such an amount of nitrogen when it is in its molten state, the investigators inferred that the hard alpha defect formed during the forging process, in which the molten titanium reacted with the contaminating oxygen and

---

[10] In metallurgy, the term "phase" refers to a given chemical composition with a distinct type of atomic bonding and arrangement of elements.

nitrogen. ([11], p.78) They further conjectured that the cavity itself was filled with hard alpha materials initially, and the brittle materials fell out during subsequent disk machining and finishing processes. ([11], p.79)

Even though the cavity was only 0.055 inches wide and 0.015 inches deep on the disk bore surface, it created a stress concentration point and initiated the fatigue crack. The fatigue crack advances incrementally with each stress cycle (i.e., each cycle of takeoff and landing in which the engine powered up and brought to operating temperature). Finally, it reached a critical size of 1.24 inches long and 0.56 inches deep at the time of the accident. ([11], p.45)

To determine how long it took between fatigue crack initiation and final failure, the investigators examined the fatigue crack surfaces using a scanning electron microscope, looking for microscopic features known as fatigue striations. Striations are marks produced on the fracture surface that show the incremental growth of a fatigue crack. Research has shown that each striation results from a single stress cycle, so the total number of striations on the crack surfaces corresponds to the component's total number of stress cycles. A standard technique for counting fatigue striations is to scan the fracture surface, using a high-powered electron microscope at 500X to 3000X magnification. When the analysts find fatigue striations in several areas, they calculate the crack growth rate across those areas, graphically integrate a plot of striation density versus distance across the entire fracture surface, and then produce a reasonably accurate estimate of the growth rate and time from initiation to final failure. ([33], p.63)

The total number of striations found on the fatigue region of the stage 1 fan disk was nearly equal to the number of takeoff-landing cycles (15,503 cycles) that the disk experienced, which indicated that the fatigue crack initiated very early in the disk's life. ([11], p.77) GEAE (General Electric Aircraft Engines) conducted fracture mechanics analysis of the fatigue crack. The analysis showed that the fatigue crack grew from a hard alpha defect area slightly larger than the size of the cavity at the fatigue origin when the fan disk was first exposed to full thrust engine power conditions. From that point, the crack growth was in line with the expected fracture mechanics behavior for Ti-6Al-4V alloy. ([11], p.77)

Finally, during the fatigue crack surface inspection, the investigators found a discolored zone roughly 0.476 inches long along the length of the disk bore and 0.180 inches deep. To understand what had caused the discoloration of the crack surface, the investigators applied secondary-ion mass spectrometry (SIMS)[11] to analyze the fatigue fracture surface's chemical composition. The analysis indicated that the chemical residues on the discolored surface correspond to compounds in the fluorescent penetrant liquid used in the fan disk's quality assurance inspections. After ruling out alternative sources of the chemical residues on the fracture surface, the NTSB concluded that a fluorescent penetrant inspection (FPI) of the fan disk caused the discoloration. ([11], p.85)

The maintenance records of the stage 1 fan disk showed that the disk had been through six inspections in its service life, each included a fluorescent penetrant inspection (FPI) of the entire disk. Which inspection was responsible for the discolored zone found on the fatigue crack surface? Using GEAE's fracture mechanics analysis, the investigators estimated the size of the fatigue crack during each maintenance inspection. The discolored area's actual size is 0.476 inches, which is closest to the estimated surface length of the crack at the last inspection (0.498 inches). ([11], p.78) Therefore, the fatigue crack surface's discoloration was created during the FPI process performed in the last inspection, which happened 14,743 (takeoff-landing) cycles since new and 760 cycles before the accident. Also, the size of the discolored area marks the crack's size at the time of the last inspection. ([11], p.85)

In sum, the fracture analysis of the stage 1 fan disk from the No.2 engine traced the origin of the fractures back to a hard alpha defect created in the disk manufacture process. At this point, the NTSB investigators understood how the stage 1 fan disk had fractured, and how the fracture and separation of the fan disk fragments had damaged all the three

---

[11] Secondary-ion mass spectrometry (SIMS) is a technique for analyzing the elemental and isotopic composition of solid surfaces. It consists of bombarding the surface with a focused primary ion beam and then collecting and analyzing ejected secondary ions.

hydraulic systems on the accident airplane. To determine how to prevent similar failures from happening again in the future, the investigators examined how the relevant design, manufacture, and maintenance practices had contributed to the failure of the No.2 engine and the three hydraulic systems.

3.5 Design and Certification Philosophies

3.5.1 Fan Disk Design and Certification

The design and certification rules for all the rotating components in a jet engine require to safeguard against metal fatigue. In the case of the stage 1 fan disk on the CF6-6 engine, the GEAE (General Electric Aircraft Engines) tests at the time of the certification demonstrated that a *defect-free* disk can sustain 54,000 takeoff-landing cycles without crack initiation. ([11], p.99) Because fatigue data have large statistical scatter and are sensitive to various manufacturing factors, the FAA only certified the engine for 54,000/3 = 18,000 cycles. The fan disk will be replaced after 18,000 cycles, regardless of any evidence of fatigue cracking. This design concept is known as “safe life” design. ([6], p.190)

However, the stage 1 fan disk from the No.2 engine of United flight 232 suffered from a fatigue failure after only 15,503 takeoff-landing cycles, which was well within the estimated safe life of 18,000 cycles. Why did the safe life design fail to protect against fatigue in this case? According to the NTSB, it was because the safe life design and certification rules assumed that the titanium alloy material used for the fan disk was free of defects.[12] Thus, the fan disk manufacturer was not required to assume that undetectable defects were present in the material when calculating the disk’s safe life. ([11], p.99)

The total number of cycles that a fan disk experiences before failure (called the fatigue life of the disk) consists of the number of cycles needed to initiate a crack and the number

---

[12] More precisely, the assumption was that the titanium alloy material that passed GEAE’s quality assurance tests and inspections during manufacture was free of defects.

of cycles needed to propagate the crack to failure. For most defect-free disks, most of the disks' total life is in the initiation of a crack, and only a minor portion of life is in the crack propagation phase. However, when a pre-existing defect is present in the material, the initiation phase of the crack growth is effectively eliminated, drastically reducing the fan disk's total fatigue life. In the United flight 232 accident, a pre-existing hard alpha defect initiated a fatigue crack very early in the disk operation. As a result, it only took 15,503 cycles before the crack grew to a critical size and caused the disk to break apart. ([11], p.99)

Based on these considerations, the NTSB concluded that the "safe life" design philosophy for critical engine components was inadequate. It further recommended an alternative "damage-tolerance" design philosophy. The main assumption of the damage-tolerance design is that the material in highly stressed areas of a component contains defects of a size just below the size detectable during manufacturing inspections. Given this assumption, the manufacturer could then use fracture mechanics to calculate the expected fatigue crack growth rate and critical crack size. The calculation is based on the stress distributions within the component and the crack propagation characteristics of the material. ([11], p.99)

The purpose of calculating the expected fatigue crack growth in the damage tolerance design is not to justify operating with a fatigue crack: If a fatigue crack is found during a normal maintenance operation, the component will be removed from service immediately and permanently. Rather, the main objective of crack growth calculation is to determine the appropriate time interval for inspections, so that a crack is more likely to be found before reaching the critical length and causing failure. ([6], p.208)

3.5.2 Hydraulic Systems Design and Certification

The NTSB investigators also identified inadequacies in the design and certification requirements of the DC-10 hydraulic systems. First, in retrospect, McDonnell-Douglas (the aircraft manufacturer) could have taken additional precautions in the design of the hydraulic systems if it had given more consideration to the potential effect of random debris from the tail-mounted engine.[13] In the DC-10, there are three different, unconnected hydraulic systems, and the hydraulic lines are physically separated to minimize vulnerability to system damage. The lines of the three hydraulic systems are widely spaced, except in the tail section, where the hydraulic lines are all lined up with the tail engine fan disk. Because of the proximity, the fan disk that flew apart on United flight 232 caused damage to all three hydraulic systems in the tail. ([6], p.177)

Why didn't McDonnell-Douglas give sufficient consideration to the potential effect of uncontained engine failure? One reason was that aircraft certification rules did not require engine manufacturers (e.g., General Electric) to provide data to aircraft manufacturers (e.g., McDonnell-Douglas) about the potential dispersion angles[14] and energy levels of fragments from engine bursts. ([11], p.90) The dispersion angle and the energy level of the fragments were calculated based on recorded observations of the results of failures both in tests and in service. Only the engine manufacturers could provide such data. Without the fragment dispersion data, aircraft designers could not accurately estimate the total interactive effect of the engine installation on the aircraft, which compromised the aircraft design's safety assessment. ([11], p.91)

Following the United 232 accident, the NTSB attempted to obtain historical and recent data regarding uncontained engine failure events but discovered no up-to-date database for such events. The most recent reports provided data only through 1983—six years earlier than the Flight 232 accident. The lack of a centrally available database of uncontained engine failures contributed to inadequate considerations of such events in design assessment and safety analysis among manufacturers. Moreover, it meant that the FAA lacked a verifiable source to research during the certification review of aircraft designs. The NTSB argued

---

[13] In all fairness, the list of possible failures considered by the engineers' design is more or less a list of historical failures: If it has never happened before, it is likely not considered unless someone thinks of it.

[14] The angle the fragments sprayed out of the engine.

that the FAA should review the current reporting requirements for aircraft manufacturers and operators, and establish a centralized database for uncontained engine failure events. ([11], p.91)

Finally, the NTSB reviewed alternative hydraulic system design concepts for wide-body airplanes to see how the DC-10 hydraulic system design could be improved. Given the three independent hydraulic systems installed on the DC-10s, the NTSB could not find a safety advantage to install another independent hydraulic system. However, the investigators observed that the existing hydraulic systems needed enhancements to protect against open or breached hydraulic lines, and to provide at least some flight control in the case of a triple hydraulic failure. ([11], p.97)

Two months after the accident, McDonnell-Douglas proposed hydraulic system enhancements that consisted of three separate installations: (1) Sensors that detect the flow rate within the hydraulic systems; (2) shutoff valves that automatically close when the hydraulic fluid within the system reservoir fall below a certain level; and (3) a cockpit warning light that alerts the pilots of the shutoff valve activation. ([11], p.67) The NTSB concluded that these enhancements could provide the flight crew with partial flight control in the event of a No. 2 engine failure similar to the UA 232 accident. ([11], p.97)

3.6 Fan Disk Manufacture and Quality Assurance

The NTSB investigators pursued three main questions about the manufacture and quality assurance inspections of Flight 232's fan disk. First, they reviewed the manufacturing process of Flight 232's fan disk to determine how the hard alpha defect and the cavity were created, and how the manufacturing process could be improved. Second, they attempted to identify the sister fan disks made from the same source material as the accident fan disk, to determine whether the sister fan disks suffered from similar defects. Finally, they examined the quality assurance inspections performed during the maintenance process to determine whether the inspections could have detected the hard alpha defect before the accident fan disk went into service.

3.6.1 The Manufacturing Process

There are three primary steps in the manufacturing of titanium alloy fan disks: Material processing, forging, and final machining. ([11], p.49) In the first step, raw materials are melted into a 28-inch-diameter, 7,000-pound titanium alloy "ingot" inside an electric arc furnace.[15] The ingot is then forged and beaten into a 16-inch-diameter "billet" with heat and massive hydraulic hammer blows. The second step involves cutting the billet into eight 700-pound pieces called "forging blanks", which are then forged into geometrical shapes close to the final shape. The third and final step involves machining the forged shape into the final fan disk shape. ([6], p.191)

Molten titanium is very reactive with air, especially with nitrogen and oxygen. When the molten titanium reacts with nitrogen and oxygen during the material processing step, it forms hard alpha defects[16] that are brittle and easily cracked. To prevent such reactions, the processing of molten titanium happens in a vacuum. Processing molten titanium at high temperature and sealing it against tiny air leaks can be very difficult and expensive. The only process used in the 1970s, and one still commonly used today, is vacuum arc remelting, or VAR. ([6], p.191)

The basic idea of VAR is to remelt the titanium alloy ingot in the same electric arc furnace after its initial melt. Hard alpha defects have a higher melting point than titanium, making them difficult to eliminate from the molten titanium. They only dissolve when held at elevated temperatures for extended periods. ([6], p.201) Before 1971, the standard industry approach to dissolving potential hard alpha defects was the double-melt VAR process, which remelted the ingot a second time in the electric arc furnace. Because of

---

[15] An electric arc furnace is a furnace using an electric arc to heat charged material.

[16] There is no chemical definition of hard alpha. Hard-alpha defects may contain up to 14.8% nitrogen and up to 2.5% oxygen.

a series of accidents caused by hard alpha defects in titanium alloy rotating engine parts, the double-melt VAR process changed in 1971 to a triple-melt VAR process, and all the fan disks manufactured after January 1972 complied with the new triple-melt requirements. ([11], p.51)

The triple-melt VAR process reduced the likelihood of hard alpha defects, but it could not guarantee a defect-free product. To determine whether the double-melt or the triple-melt VAR process produced Flight 232's fractured fan disk, the investigators used the serial number of the accident fan disk to trace the titanium alloy ingot from which the fan disk was made. Manufacture records showed that the source ingot was melted on February 23, 1971, shortly before General Electric changed its material specification to require triple-vacuum melting. Moreover, the timing of the change was such that the accident fan disk was among the last fan disks produced using the double-melt process. ([11], p.51) In response to this discovery, the FAA mandated inspections of all the fan disks manufactured using the superseded double-melt VAR process, and General Electric announced plans to replace all of them within 1,500 flight cycles. ([6],201)

The NTSB investigators further examined other improvements in the titanium manufacturing process since 1971. In addition to the triple-melt process, the industry also introduced improvements, including minimizing air leaks, tighter quality control on the raw materials, and improved furnace cleaning requirements to reduce contamination. The investigators concluded that the current (i.e., 1989) technology for titanium manufacturing had progressed to the point where critical defects were rare. Additional reductions in the number and size of hard alpha defects were unlikely to occur without fundamentally changing the production process. ([11], p.78)

Finally, the investigators examined when and how the cavity at the fatigue origin formed during the manufacturing process. Based on features of the mechanical deformation on the cavity bottom, and based on the location and orientation of the microcracks beneath the cavity surface, the investigators argued that the cavity was most likely created by shot peening, a cold work process in the final machining step of fan disk manufacturing. Shot peening involves blasting the finished fan disk surface with small spherical shots, which normally strengthens the metal surface and increases its ability to resist fatigue fracture. In the case of Flight 232's fan disk, however, the final machining and shot peening removed the hard alpha material that originally filled the cavity, causing a stress concentration that initiated the fatigue crack. ([11], p.79)

3.6.2 Tracing the Sister Fan Disks

The accident fan disk was one of eight fan disks made from the same titanium alloy billet. The fact that it had a hard alpha defect immediately raised the question about whether its sister fan disks had similar defects. After the Flight 232 accident, the investigators moved quickly to trace the sister fan disks and recall them for immediate inspection. ([6], p.200)

Tracing the sister disks of the accident fan disk was no minor task. First, the fan disks in question were created in 1971, 18 years before the accident and the investigation. Second, the three main steps of the fan disk manufacturing process were carried out by different companies. The final machining of the accident fan disk was done by General Electric Aircraft Engines (GEAE), whose records showed that the disk was forged by Aluminum Company of America (ALCOA). In turn, ALCOA had subcontracted with suppliers such as Titanium Metals Corporation of America (TIMET) and Reactive Metals Incorporated (RMI) to supply raw materials in billet forms. ([11], p.51) Finally, the records obtained from the companies contained gaps, errors, and contradictions, further complicating the traceability of the fan disks. ([11], p.80)

Every CF6-6 fan disk had a serial number for unique identification, and a heat[17] number (given by the titanium supplier) for tracing the billet from which the disk was made. The serial number for the accident fan disk was MPO 00385. ([11], p.80) ALCOA records showed

---

[17] In metallurgy, a"heat" refers to the batch of raw materials that were melted to produce a titanium alloy billet.

that fan disk MPO 00385 originated from a billet with the heat number K8283 created by TIMET on February 23, 1971. ALCOA records further showed that this billet produced eight forging blanks, corresponding to fan disks with serial numbers MPO 00381 through 00388. ([11], p.82)

The investigators then cross-checked the GEAE records, and found that six of the seven sister fan disks (MPO 00382, 00383, 00384, 00386, 00387 and 00388) were in service at the time of the accident. GEAE recalled these six fan disks for testing, and found that fan disk MPO 00382 and MPO 00388 also had rejectable hard alpha defects. ([11], p.48)

The GEAE records, however, contained several anomalies. First, GEAE did not have any records for fan disk MPO 00381. Instead, the GEAE records mentioned another fan disk with the same serial number (i.e., MPO 00385) as the accident fan disk. However, there was no evidence at ALCOA that the forging company had shipped two disks with serial number MPO 00385 to GEAE for machining. ([11], p.82) To distinguish between the two MPO 00385 fan disks on GEAE's records, the investigators labeled the accident fan disk "disk B", and the other fan disk "disk A". ([11], p.53)

Second, unlike the accident fan disk (disk B), which had the heat number K8283, disk A had a different heat number 704233 (according to GEAE records). K8283 was a heat number used by the titanium supplier TIMET to denote one of its titanium alloy billets. In contrast, 704233 was a heat number used by a different titanium supplier RMI. If both heat numbers were correct, it follows that disk A and the accident disk (disk B) came from two different titanium alloy billets made by two different companies. In that case, they were not sister disks after all. ([11], p.80)

Finally, ALCOA records and GEAE records disagreed on the whereabouts of forgings made from billet 704233. According to the ALCOA records, billet 704233 were not made into fan disks at all. Rather, all the forgings made from billet 704233 were for airframe parts, and none of the forgings were delivered to GEAE. ([11], p.55) In contrast, the GEAE records stated that one and only one forging made from billet 704233—namely fan disk A—was processed by GEAE. Moreover, GEAE records indicated that disk A was rejected after an unsatisfactory ultrasonic inspection on June 7, 1971. Disk A was reportedly scrapped and cut up for examination on November 1, 1972, but no evidence of a hard alpha or other defect was found. ([11], p.53)

The contradictions between the ALCOA records and the GEAE records raised a host of questions. On the one hand, if the ALCOA records were correct, what happened to the sister disk MPO 00381? If disk MPO 00381 were the same disk as disk A on the GEAE records, why did GEAE associate disk A with a different heat number 704233? On the other hand, if the ALCOA records contained errors, then what were the sister disks of the accident fan disk? Which titanium supplier provided the source material from which the accident disk and its sister disks were made?

Lacking independent records to resolve the contradictions, the investigators decided to perform chemical analysis on the accident fan disk (MPO 00385) and its six purported sister disks (MPO 00382, 00383, 00384, 00386, 00387 and 00388), in an attempt to verify their relationships to each other and their source material. After taking multiple samples from the bore and the rim of each of the seven disks, the investigators coded the samples before distributing them to GEAE, ALOCA, TIMET, and RMI for an independent analysis. ([11], p.80)

Unfortunately, the combined results of the four analyses were inconclusive and raised even more questions. For instance, the titanium supplier TIMET argued that certain trace elements that should have been present in any TIMET product were not detected in sufficient quantity in some of the disks. TIMET's and GEAE's chemical analysis divided the seven disks into two groups based on variations in trace chemical elements: Disks MPO 00383, 00384, and 00387 belonged to one group, whereas disks MPO 00382, 00385, 00386 and 00388 belonged to another. These two analyses further stated that TIMET produced only the first group. ([11], p.57) In contrast, the titanium supplier RMI argued that (1) all seven disks could be from the same billet, and the variations in chemical elements were normal, (2) none of the seven disks could have been produced from the billet 704233 created by RMI. Given such conflicting analysis, the NTSB investigators were unable to determine if the seven disks came from the same billet (and if so, which titanium supplier produced it) or from different billets. ([11], p.80)

However, if the seven fan disks were not produced from the same billet (heat number K8283) as ALCOA and GEAE records suggested, then the records on a large number of GEAE fan disks were suspect. Because of doubts about the records, the FAA could not determine whether all the sister disks of the accident disk had been removed from service. Also, suppose the records on heat numbers (i.e., disk origins) were not accurate. In that case, some double-melted disks might be mistakenly identified as triple-melted disks, making the FAA mandate on inspecting all double-melted disks ineffective. ([11], p.81)

In the end, the puzzles about the records on the accident fan disk and its sister disks remained unresolved. Consequently, the NTSB concluded that the manufacturers' recordkeeping programs on the manufacture of CF6-6 fan disks in the early 1970s were deficient. It further recommended that the FAA conduct a comprehensive evaluation of the manufacturing recordkeeping practices, in order to ensure traceability of critical airplane parts in the event of in-service failures. ([11], p.83)

3.6.3 Quality Assurance Inspections

During the manufacturing process for the accident fan disk, the source material of the disk (i.e., the titanium alloy billet) and the disk part went through four nondestructive, quality assurance inspections. These inspections were supposed to detect anomalies, either internally or on the disk surface. ([11], p.57) The NTSB investigators reviewed each of the inspection processes to determine whether any of them could have detected the hard alpha detect that initiated the fatigue crack. ([11], p.83)

The first inspection occurred after the formation of the titanium alloy billet from which the accident fan disk was made. It was an ultrasonic inspection of the 16-inch diameter billet with heat number K8283, carried out by the titanium supplier TIMET.[18] ([11], p.58) The ultrasonic inspection used a special piezoelectric[19] crystal that could convert electrical energy into mechanical energy, and vice versa. The process works as follows: When receiving an electrical signal, the crystal vibrates and generates high-frequency sound waves that propagate through the billet. When the sound waves encounter an internal discontinuity (such as a crack inside the billet), some energy is reflected from the flaw surface. The reflected wave signal vibrates the crystal and generates a measurable return signal. Information about the location, size, orientation, and other features of the flaw surface can sometimes be inferred from the return signal. ([6],195)

However, ultrasonic detection of hard-alpha defects in a titanium alloy billet can be difficult. First, the hard alpha caused by excess nitrogen and oxygen is only slightly different in composition from the surrounding material; and the smaller nitrogen and oxygen atoms between the titanium atoms have little effect on the base material's ultrasonic properties. ([6], p.195) Second, information from the titanium industry indicated that nearly all the hard alpha defects that had been detected ultrasonically were associated with relatively large cavities. However, certain hard alpha defects may not be associated with large cavities, at least not during the billet stage of the manufacturing process. ([11], p.84) The cavity on the bore surface of the accident fan disk, for instance, was most likely created during the final machining step of the manufacturing process. It probably did not exist during the initial material processing step (e.g., billet formation), when the hard alpha defect was still buried inside the billet[20] and likely filled with material.

Consequently, the investigators concluded that the first ultrasonic inspection probably could not have detected the hard alpha defect in the source material of the accident fan disk.

---

[18] Here we take the ALCOA records about the origin of the accident fan disk for granted, and put aside the worries about the reliability of the records.

[19] Piezoelectric means the ability of a material to create an electric charge in response to mechanical stress.

[20] A billet has a greater diameter than the fan disk shapes cut from it, so defects that were inside the billet could be on the surface of a fan disk cut from it.

Nevertheless, the investigators noted the billet inspection process had been improved in an important aspect since the accident. Namely, the size of the billet was reduced from 16-inch diameter to 10-inch diameter. This smaller diameter allows a more sensitive ultrasonic inspection of the billet. Moreover, the increased hot working required to reduce a 28-inch diameter titanium ingot to 10 inches instead of 16 inches increases the likelihood of any hard-alpha defect becoming a crack or void. This, in turn, increased the likelihood that defect can be detected during subsequent ultrasonic inspections. ([6], p.195)

The next three inspections were all performed by General Electric Aircraft Engines (GEAE) in 1971, during the final machining step of the manufacturing process. They consisted of an ultrasonic inspection and a macroetch inspection on a preliminary disk shape called the rectilinear machined forged shape (rectilinear shape for short), and a fluorescent penetrant inspection (FPI) on the final disk shape. ([11], p.58)

The second inspection was an ultrasonic inspection of the disk forging after GEAE had machined it to the rectilinear machine forged shape. The rectilinear shape is very close to the final machined shape. For instance, on the disk's bore surface, only about 0.015 inches is removed from the rectilinear shape during machining to the final disk shape. However, the investigators argued that the hard alpha defect on the rectilinear shape likely were not detectable by the ultrasonic inspection. This was because the ultrasonic inspection on the rectilinear shape could have detected the hard alpha area only if there had been cracking or voids associated with the defect, and the accident fan disk *in its rectilinear shape* probably did not have cracks and voids associated with its hard alpha defect. ([11], p.83)

First, the hard alpha area on the accident fan disk had a cavity at its center. However, the cavity was most likely caused by the shot peening process, which occurred after the rectilinear shape had been machined into the final disk shape. Therefore, it is likely that the hard alpha area did not have a cavity associated with it when the accident fan disk was still in its rectilinear shape. ([11], p.83)

Second, the hard alpha area on the accident fan disk also contained microcracks parallel to the fatigue crack that caused the disk fracture. However, these microcracks were likely introduced into the disk after the ultrasonic inspection of the rectilinear shape. The main evidence for this claim came from a (purported) sister disk of the accident fan disk, namely disk MPO 00388. An ultrasonic inspection of disk MPO 00388 after the accident detected hard alpha areas. Metallurgical evaluation of the ultrasonically located defect in MPO 00388 showed significant microcracks in the hard alpha areas; these cracks led to the ultrasonic detection of the defects. However, disk MPO 00388 was also ultrasonically inspected during 1971 when it was still in the rectilinear shape, but no indications of a rejectable hard alpha area were reported at the time. This fact suggested that the microcracks associated with the defects in MPO 00388 were introduced after the 1971 ultrasonic inspection of the rectilinear shape. A similar argument can be made concerning the hard-alpha-related microcracks in the accident fan disk. ([11], p.84)

The third inspection was a macroetch inspection by GEAE, also on the rectilinear machined forged shape. Macroetching involves etching the surface of a specimen with a suitable acid or reagent so that anomalies on the surface can be visible with the unaided eye. The accident fan disk was macroetched while in the rectilinear shape, but not in its final shape. GEAE stated that the fan disk's final shape was not macroetch inspected for various reasons, including concern that the etching procedure would remove too much of the surface material. ([11], p.84)

Unfortunately, the choice of performing the macroetch inspection on the rectilinear shape rather than on the final shape of the fan disk likely made a difference to the detectability of the hard alpha defect. Unlike ultrasonic inspections, macroetch inspections can only detect anomalies on the surface of a specimen. In the case of the accident fan disk, the hard alpha defect was on the surface of the final machined shape, but it was still a fraction of an inch below the surface on the rectilinear shape. Postaccident tests performed by the NTSB investigators showed that neither the macroetch inspection procedure used by GEAE nor the procedures used by other major engine manufacturers could have detected a subsurface defect. ([11], p.84)

Given that the cavity on the accident fan disk bore was most likely created during machining to the final disk shape (i.e., shot peening), the investigators further concluded that the cavity would have been apparent if the accident disk had been macroetched in its final shape. ([11], p.85) Consequently, the NTSB recommended that the FAA require engine manufacturers to perform a surface macroetch inspection of the final shape of critical titanium alloy rotating components during the manufacturing process. ([11], p.106)

The fourth and last inspection of the accident fan disk before it entered service was a fluorescent penetrant inspection (FPI), performed on the final disk shape by GEAE on December 9, 1971. Like macroetching, FPI is also a surface inspection technique, and it works via a phenomenon known as capillary action. Capillary action is a liquid flow in narrow spaces without the help of, or even contrary to, external forces like gravity. During FPI, a liquid penetrant is applied to the inspected surface, so that it can flow with capillary action into any surface defects. After some time for the penetrant to soak into potential defects, the inspector wipes the excess penetrant from the surface, then applies a developer (typically a white powder) to the surface. The developer draws the penetrant out of defects by reverse capillary action. The penetrant bleeds into the developer to produce color indications more visible than the defects themselves. ([6],197)

The FPI performed on the final disk shape in December 1971 did not find any anomalies. ([11], p.58) Should the inspection have detected the hard alpha defect or the cavity on the bore of the accident fan disk? The NTSB investigators did not address this question directly. Part of the reason for this was that the accident disk went through six more FPIs inspections during its service life. None of the subsequent FPIs detected the fatigue crack that originated from the hard alpha origin, which eventually became much larger than the hard alpha area itself. Instead of focusing on the first FPI performed at the end of the manufacturing process, the investigators turned their attention to the six subsequent FPIs performed during the accident disk's service life.

3.7 Maintenance Inspection Procedure

The accident fan disk's maintenance records showed that the disk had been through six detailed maintenance inspections in its lifetime, each of which included a fluorescent penetrant inspection (FPI) of the entire disk performed by United Airlines' maintenance personnel. All six inspections had been stamped and accepted by the inspectors with no crack observed. The inspection records appeared to comply with United Airlines' maintenance program approved by the FAA. ([11], p.85) However, it was also true that the fatigue crack that eventually broke apart the disk had been growing steadily throughout the 17-years service life of the disk, and none of the FPIs had detected it. In particular, the last FPI was done only about a year (or 760 flight cycles) before the accident. Could, or should, any of the FPIs—especially the last one—have detected the fatigue crack?

Based on the GEAE fracture mechanics analysis, the investigators estimated the size of the fatigue crack during each inspection. The surface length of the fatigue crack at the last inspection was estimated to be 0.498 inches. Moreover, the investigators found a discolored zone roughly 0.476 inches long along the crack surface after the accident. This size corresponds reasonably well with the predicted size of the crack at the last inspection. The examination of the chemical composition of the discolored surface found residues corresponding to compounds in the fluorescent penetrant liquid used in the inspections. Therefore, the investigators concluded that the discolored area marked the crack size at the time of the last inspection, and that the crack was sufficiently open that the FPI fluid entered the crack. ([11], p.85)

Unlike the cavity inside the hard alpha region, which was only 0.055 inches long and could be difficult to detect, a fatigue crack about 1/2 inch long along the bore surface should have a high probability of detection if a proper FPI was conducted. ([11], p.86) The GEAE fracture mechanics analysis indicated that the crack sizes during several inspections before the last inspection would also have been detectable by FPI under normal conditions.

However, the crack was not detected, and consequently, the fan disk was considered to be free of defects and was accepted as a serviceable part. ([11], p.87) This raised the question: Why didn't the inspectors detect the fatigue crack during those inspections when the crack should have been detectable?

A review of the FPI inspection process suggested a variety of possible explanations for the inspector's failure to detect the crack. First, the fan disk was suspended by a cable during the FPI. The inspectors might not adequately rotate the disk to prepare and view all portions of the disk bore, particularly the areas hidden by the suspension cable. Second, it is possible that loose developer powder had dropped on top of the crack and obscured it sufficient to prevent detection. Finally, inspection experience indicated that due to the disk's geometry, certain disk parts frequently showed clear FPI indications and that other areas rarely did so. For instance, the perimeter of the disk web was an area of frequent FPI indications, whereas the central disk bore area rarely produced FPI indications. Hence, it is possible that the inspector did not consider the bore area a critical area for inspection, and gave it only cursory attention. Any of the above possibilities could have contributed to the non-detection of the fatigue crack in the accident fan disk. ([11], p.87)

Because of these discoveries, the NTSB investigators were concerned that manual inspections in general, and FPI in particular, might be susceptible to human factors problems that could significantly degrade inspector performance. Moreover, there was a minimum redundancy built into the aviation industry's manual inspection processes to prevent human error. For instance, inspectors typically work independently and receive very little supervision. ([11], p.87) Consequently, the NTSB concluded that the FAA should develop a research program to identify emerging technologies that could simplify or automate the inspection processes. Moreover, the FAA should also encourage the development of redundant (i.e., "the second pair of eyes") inspection oversight for critical part inspections, such as rotating engine part inspections. ([11], p.88)

3.8 Flight Performance and Emergency Management

Finally, the NTSB investigators evaluated the performance of the United 232 flight crew and the emergency response on the ground to determine whether anything could be improved in the future.

Given the unprecedented triple hydraulic failures, the fact that the flight crew of United 232 was able to land the aircraft at all was remarkable. Still, the investigators wanted to know whether DC-10 pilots could be trained to control the airplane and land safely with no hydraulic power to activate the flight controls. So the NTSB directed a simulator reenactment of the events leading to the United flight 232 crash. The exercise simulated the failure of the No.2 engine and the loss of fluid in all three hydraulic systems. ([11], p.72) Forty-five simulated flights were flown by qualified DC-10 line captains, training captains, and test pilots. The only means of control for the test pilots were the wing engines, just like in the accident. The comments, observations, and performance of the pilots were then recorded and analyzed. ([6], p.173)

The simulator tests showed that such a maneuver involved too many unknown variables. Landing at a predetermined point and airspeed on a runway was a highly random event. The lack of controllability during the approach and landing made it virtually impossible to design an effective simulator training exercise. Consequently, the NTSB concluded that the plane could not be safely landed even with additional training. The NTSB further stated that "under the circumstance, the UAL (United Airlines) flight crew performance was highly commendable and greatly exceeded reasonable expectations." ([6], p.173)

Concerning the emergency response at the Sioux Gateway Airport, the investigators believed that the management was effective overall. The established airport and county emergency plan, the recent full-scale disaster drill in 1987, and the warning time given by the United 232 flight crew contributed to efficient emergency preparation and management near the crash site. ([11], p.93)

Nevertheless, the investigators did notice two problems that the airport firefighters had with controlling the post-crash fire. First, the main cabin of the accident airplane ended up in a cornfield at the end of the crash landing runway, and the height and density of the cornstalks interfered with the firefighters' ability to control the spread of fire within the cabin. The cornstalks also made it difficult for the firefighters to see the passengers, and for the passengers to find a path away from the burning cabin. Therefore, the investigators argued that crops that limit visibility and mobility should not be cultivated on certified airports. ([11], p.94)

Second, the water pump on a fire truck failed during the resupply attempts, and as a result, no extinguishing agent was applied to the fuselage for about 10 minutes. During this time, the fire intensified and spread inside the main cabin. The examination of the failed pump revealed a problem with the design of the suction hose assembly, which caused the suction hose to collapse and block the water flow. ([11], p.94) The investigators further found that similar problems had occurred before at the U.S Air Force Base, but the Air Force did not take immediate action to correct the problem. In particular, there were no requirements to routinely test all fire service equipment at their full discharge capacity, which allowed equipment deficiencies to go unnoticed until an emergency occurred. Consequently, the NTSB concluded that the FAA and the Air Force should require routine testings of all fire service equipment at their full capacities. ([11], p.95)

3.9 Conclusion

According to the NTSB, the "probable cause" of the United flight 232 accident was inadequate consideration of human factor limitations in the maintenance inspection procedures. The inspections failed to detect a fatigue crack originated from a previously undetected hard alpha defect in the bore of a stage 1 fan disk. The fan disk suffered from a catastrophic disintegration during the accident flight, scattering debris that severed the lines of all hydraulic lines, rendering the accident aircraft uncontrollable. ([11], p.102)

Based on the deficiencies and vulnerabilities discovered in the accident investigation, the NTSB made numerous safety recommendations to the FAA. First, concerning the design and certification requirements for the DC-10, the recommendations included: adopting the damage-tolerance design philosophy for critical engine rotating parts, establishing a centralized and update-to-date database for uncontained engine failure events, and researching hydraulic system enhancements for minimal flight control in the event of a triple hydraulic systems breach. ([11], p.102-103)

Second, concerning the fan disk manufacturing and maintenance procedures, the NTSB recommended a comprehensive evaluation of the manufacturers' recordkeeping programs to ensure the traceability of critical airplane parts. It further recommended developing techniques for simplifying and automating inspections of critical parts and practices that provide redundant inspection oversight (i.e., "second pair of eyes"). ([11], p.103-104)

Finally, concerning emergency preparation and management on the ground, the NTSB recommended that airport certification include examinations of airfield terrain to ensure surface obstructions such as crops do not interfere with rescue and firefighting activities. Also, the airport operators should regularly perform maximum capacity tests of all the emergency response equipment and vehicles. ([11], p.107)

Chapter 4

Case Study 3: United Flight 585

4.1 History of the Flight

On March 3, 1991, United Airlines Flight 585, a Boeing 737-200 aircraft, crashed near Colorado Springs Municipal Airport in Colorado Springs, Colorado. All 25 people on board died from the crash. ([12], p. vi)

Flight 585 was a scheduled flight from Denver, Colorado, to Colorado Spring, Colorado. It departed from Denver at 09:23 (Mountain Standard Time) and was scheduled to arrive at Colorado Springs at 09:46. The flight crew included the captain, first officer, and three flight attendants. Because it is a short trip, Flight 585 stayed below 11,000 feet above the sea level. According to the weather message received by the flight crew, Colorado Springs had clear weather, great visibility, and strong wind (about 23 knots[1] with gusts[2] up to 33 knots). ([12], p.2) As the flight crew began their approach to the airport, they received a weather update. It added that low altitude wind shear[3] warnings were in effect at Colorado Springs Municipal Airport. ([18], p.25)

---

[1] A knot (abbreviated as kts) is a unit of speed measurement, with 1 knot equal to 1 nautical mile per hour, or 1.15 miles per hour.

[2] A gust is a sudden increase of the wind's speed that lasts no more than 20 seconds.

[3] Wind shear is a change in wind direction or speed over a short distance.

At 09:37, the airplane descended to 8,500 feet. The first officer contacted the Colorado Springs control tower, which cleared United flight 585 to land on runway 35. ([12], p.3) In the radar transmission, the first officer asked the tower if any other aircraft reported gains or losses of airspeed, which would be signs of wind shear. The tower responded with reports from another 737, which lost 15 knots at 500 feet, gained 15 knots at 400 feet, and gained 20 knots at 150 feet. ([18], p.26)

At 09:41:23, the tower directed the flight to hold short of runway 30 after landing, and the first officer confirmed the message. This transmission was the last one received from flight 585. ([12], p.4)

Meanwhile, many witnesses observed that the airplane was flying at an altitude lower than normal. Despite this, it appeared to be operating normally, until it suddenly rolled to the right and pitched nose down. ([12], p.4) Soon the airplane entered a near-vertical dive, and hit the ground in an area known as Widefield Park, 2.47 nautical miles south of the south end of runway 35 at the Colorado Springs Municipal Airport. The time of the impact was about 09:43:43 Mountain Standard Time. ([12], p.4, [18], p.15)

4.2 The Structure of the NTSB Investigation

Within hours after the crash of Flight 585, the National Transportation Safety Board (NTSB) formed a “Go Team” consisting of five members of the board. The purpose of the Go Team was to begin the investigation of the crash at the crash site as quickly as possible, and to assemble a board spectrum of technical expertise needed for the investigation. ([16]) By the nightfall of the day, the Go Team held the first official meeting of the investigation in Colorado Springs. Within days, a total of 18 NTSB investigators were assigned to the Flight 585 case. They were joined by specialists from other organizations participating in the investigation (called “parties” in the NTSB jargon), including Boeing, United Airlines, the Federal Aviation Administration (FAA), and the Air Line Pilots Association (ALPA). These experts were then assigned to their appropriate groups on the investigative team. ([18], p.18-19)

By the day after the crash, both of the black boxes—the flight data recorder (FDR) and the cockpit voice recorder (CVR) were recovered from the crash site and sent to the NTSB's laboratories in Washington, D.C. for readout. ([12], 37) In the meantime, other working groups had already started investigations in their areas of expertise. Three of the groups—Structures, Powerplants, and Systems—began to examine the wreckage at the crash site. ([18], p.31)

Because Flight 585 crashed nose-first at over 200 knots, the wreckage was compressed entirely into a crater only 9 feet deep, 24 feet wide and 39 feet long. As a result, wreckage recovery required heavy construction equipment such as backhoes and cranes to dismantle and extract wreckage piece by piece. ([18], p.29) In order not to miss possible clues that might be disturbed or destroyed in this recovery process, the Structures, Powerplants, and Systems Groups had to document the aircraft parts and their relative positions when removing them from the crater. Once the pieces were labeled and photographed, they were shipped to a hanger in Colorado Springs for further examination. ([18], p.34)

The responsibility of the Structures Group was to document the airframe wreckage and the accident scene. This group looked for problems with the aircraft's overall structure and fuselage that could have contributed to the accident. In addition, the group would calculate the plane's impact angle to help determine the plane's pre-impact course and altitude. ([16])

The Powerplant group, in contrast, focused on examining the remains of the engines, the propellers, and engine accessories. The fundamental question for this group was whether the engines were working properly at the moment of ground impact, and if not, whether the engine malfunctions had contributed to the crash. ([18], p.31)

Finally, the Systems Group sought to determine whether any component of the plane's hydraulic and electric systems[4] was a contributing factor. In this case, the Systems group focused on examining the flight control surfaces—movable surfaces on the wings (such as flaps and ailerons) and the tails (such as the elevators and the rudder)—and the complex machinery that controls the motions of these surfaces. ([16])

While the Structures, Powerplants, and Systems Groups were busy examining the wreckage, a fourth group—Operations—was charged with establishing the background of the aircraft and its last flight. Unlike the groups that dealt mostly with post-crash wreckage, the Operations Group had a more historical perspective. It consisted of more specialized subgroups ([18], p.34-36):

- The Air Traffic Control (ATC) Group determined the air traffic services given to the airplane, based on ATC radar data and transcripts of radar transmissions between the controller and the pilot.

- The Performance Group reconstructed the history of flight 585's final flight, including details of its flight path in the final minutes before the crash, based on the flight recorder and radar data, witness observations, and simulation studies.

- The Maintenance Group looked into the maintenance records of the accident airplane, trying to find anomalies in previous flights that might offer clues to the accident.

- The Human Performance Group probed into the personal and medical histories of the two pilots of Flight 585, searching for anything that might have contributed to the accident.

- The Meteorology Group collected all pertinent weather data from the National Weather Service and local weather stations for a large area around the accident scene. The main goal was to reconstruct a picture of the weather during the accident.

---

[4] A hydraulic system uses pressurized fluid to transmit and control energy.

Finally, the Survival Factors Group documented deaths and injuries and evaluated the evacuation, rescue, and community emergency planning. ([16]) Since no one survived the Flight 585 crash, this group was the first to complete its investigative work.

4.3 The Black Boxes

The NTSB's recorder specialists in Washington D.C. examined Flight 585's Flight Data Recorder (FDR) and Cockpit Voice Recorder (CVR) for details for its final flight. Both recorders had sustained great damage from the impact of the crash, although the tapes were largely intact, and the investigators were able to extract the data in them. Unfortunately, even though FDR could record many flight parameters, it was set up in the accident airplane to record just five parameters: heading[5]; altitude[6]; airspeed[7]; vertical acceleration measured in G loads[8]; and microphone keying[9]. Each parameter was sampled and measured once every second; except for vertical acceleration, which was sampled 8 times per second. ([12], p.37) The CVR on board of 585 was designed to record the flight 30 minutes of a flight. Since the accident flight only lasted about 20 minutes, the CVR recorded nearly all the cockpit conversations during the final flight.

To assemble a picture of the accident flight's final minutes, the investigators from the Performance Group correlated the FDR data, the CVR data, and radar of data showing Flight 585's last moments in the air. The picture looks roughly like the following:

At 09:42:29, about 70 seconds before impact, the first officer announced: "10 knot change

---

[5] The heading of an aircraft is the compass direction the aircraft's nose. It is typically expressed relative to due north on a compass and measured clockwise. For instance, north is 360 degrees, east is 90 degrees, etc.

[6] Altitude is measured relative to the sea level.

[7] Airspeed is the speed of an aircraft relative to air.

[8] G-load is a unit of acceleration that equals the acceleration of Earth's gravity. The force on an accelerating body can be expressed as a multiple of Earth's acceleration.

[9] Keying a microphone is simply holding down the mic button without saying anything. The FDR records the microphone keying as a binary value (1 is keying, 0 is not keying) at a given time. Since microphone keying are common events recorded by both the FDR and CVR, they can be used to align the two time-series on the two recorders.

there." The captain replied: "Yeah, I know...an awful lot of power to hold that airspeed." By that time, the airspeed was about 155 knots with 2-10 knot variations; the airplane was approaching the runway 35 at a heading of 300 degrees;[10] the normal vertical acceleration varied between 0.6 and 1.3 G; and the indicated altitude was 8,000 feet[11]. The flight was still routine at this point. ([12], p.5)

At 09:42:50, about 50 seconds before the impact, the aircraft began a normal descent. 10 seconds later, it began to deviate from steady flight by descending at about 2,200 feet per minute, a faster rate than required by a standard approach to the runway. At the same time, its heading began to change 0.5 degrees per second towards the north until it was about 320 degrees. ([12], p.5)

At 09:43:01, about 40 seconds before impact, the first officer commented: "Another 10 knot again." Two seconds later, the captain called for "30 flaps". The "30 flaps" command allows a reduction of airspeed to prepare for landing; at the time, the airplane's indicated airspeed was about 160 knots. ([12], p.5)

At 09:43:08, about 33 seconds before impact, the first officer emitted a "Wow", but there was no additional explanation or comment from either pilot for the next 20 seconds. ([18], p.27)

At 09:43:20, about 20 seconds before impact, the rate of heading change to the right increased, but it was still consistent with a 20-degree back angle right turn to align with the runway. ([12], p.5)

At 09:43:28, about 13 seconds before impact, the first officer called out, "we are at a thousand feet". Two seconds later, another heading change to the right began and continued at about 4.7 degrees per second, nearly twice the rate of a standard turn. ([12], p.5)

At 09:43:33, 8 seconds before impact, the first officer said, "Oh God, flip..." when the

---

[10] Since north is 360 degrees and west is 280 degrees, 300 degrees means that the nose of the airplane was pointing to the west of a northwest direction.

[11] The indicated altitude is relative to the sea level. The Colorado Springs Municipal Airport is at 6,187 feet above the sea level.

captain called for “15 flaps”. The word “flip” was not spoken clearly by the first officer, and was partially interrupted by the captain’s flap command. ([18], p.27) The selection of the 15 flaps is consistent with the initiation of a go-around.[12] ([12], p.5)

In the last 7 seconds of the flight, the altitude decreased rapidly; the vertical acceleration increased from 0.7 G to 4.09 G; the airspeed increased drastically from 160 knots to its final recorded value of 213 knots; the heading swung sharply to the south until it was pointing almost in the opposite direction. ([18], p.25) At 09:43:41, the sound of impact marked the end of the CVR recording. ([12], p.39.)

This picture of the last minute of the accident flight is consistent with the witness reports. According to witness observations, the airplane was banking right in its final approach alignment with the runway. When it was about to align with the runway, the aircraft momentarily leveled the wings and then began to roll to the right at a steady rate. The roll continued until the airplane flipped over with its nose nearly straight down. ([18], p.5) However, nothing in this picture indicates *why* the accident airplane fell out of the sky and crashed in this way.

Furthermore, the flight crew comments on the CVR suggested that they were caught by surprise by some rapidly developing event, during which the airplane lost control. Furthermore, they were alert and active during the final 9 seconds, and they almost certainly tried to counteract the roll of the airplane with control wheel rotation. ([12], p.80) If so, why couldn’t the flight crew prevent the loss of control of flight 585? Assuming that the pilots applied the control countermeasures promptly, what events could have produced rolling motions of the airplane that could not be countered by the pilot inputs? This question would become the focus of the Flight 585 investigation.

---

[12] A go-around of an aircraft is an aborted landing.

4.4 Searching for Clues

To explain the loss of control of flight 585 and the inability of the flight crew to regain control, the NTSB investigators considered a variety of possible scenarios, including human errors, structural failures, engine failures, malfunctioning flight control systems, and atmospheric disturbances. ([12], p.80)

Over the weeks after the crash, as the different NTSB groups updated their progress, many candidate scenarios were ruled out. The Human Performance Group found nothing in the flight crew's personal or medical histories relevant to the accident. The Air Traffic Control Group reported that the controllers had not been at fault in their work. ([18], p.36) The Structures Group found that the wreckage was localized at the crash site, meaning that no pieces of the airplane had separated before the crash. Fracture analysis of the wreckage failed to produce any evidence of pre-impact structural problems. ([12], p.83)

The Powerplant Group considered the possibility that one or both engines failed during the final minute of the flight and caused the loss of control. To evaluate this possibility, the Powerplant investigators carefully examined the interior of the engines for traces of explosion or fire damage. However, all signs pointed to the engine and turbine blades rotating normally at impact. ([18], p.32). Furthermore, they recovered all the engine indicator dials from the cockpit instrument panel. They found indentation made by the indicator needle on the face of the dial at the moment of impact. These indications showed that both engines were producing nearly equal thrust at impact. ((([18], p.33) Finally, the investigators also analyzed the spectrum of frequencies recorded in the CVR, and found signatures consistent with the engines' characteristic frequencies in the last 15 seconds of the flight. ([12], p.81) Based on these discoveries, the investigators concluded that engine failures were not factors of the accident.

As more and more possible scenarios of the accident were ruled out, two significant leads worthy of further investigation emerged. One relates to the accident aircraft's rudder control system, and the other relates to weather events near the Colorado Springs Municipal Airport.

The Systems Group of NTSB carefully examined all recovered components of the flight control system to find any anomaly that could have contributed to the loss of control. ([12], p.84) The flight control system on the 737 controls the rotation of the aircraft in three dimensions: ([9], p.11-12.)

1. First, rotation around the wing-to-wing axis is called **pitch**. Pitch is primarily controlled by the elevators on the horizontal tail surface. When the pilot moves the control column forward or aft, the elevators move up or down, allowing the airplane's nose to move up or down.

2. Second, rotation around the nose-to-tail axis is called **roll**. Roll is primarily controlled by the ailerons on the outer rear edge of each wing. When the pilot turns the control column left or right, the two ailerons on the two wings move in opposite directions (up or down), allowing the airplane to roll to the left or right.

3. Third, flight control about the vertical axis is called **yaw**. Yaw is primarily controlled by the rudder on the vertical tail fin. When the pilot pushes the right or left rudder pedal forward or aft, the rudder moves side to side, causing the tail to move left or right corresponding to the pilot input. As a result, the nose of the airplane yaws left or right.

From the FDR data, it was clear that Flight 585's deviation from steady flight began with a roll to the right, and both the roll control system malfunctions and the yaw control system malfunctions could have produced such a maneuver. A pitch control system malfunction, on the other hand, would have produced a sudden change in vertical acceleration, which was not obvious from the FDR data. In addition, there were no anomalies in the recovered elevators and other pitch control components. The investigators thus concluded that the pitch control systems did not play a role in the accident. ([12], p.85)

In addition, the System investigators examined the recovered aileron power control units and found that the ailerons were at or near neutral positions at impact. Examination of other auxiliary roll control systems also failed to uncover any malfunction capable of producing an uncommanded roll movement. ([12], p.85)

Examination of the rudder control system, however, raised concerns. The first concerning discovery came from the Maintenance Group, which found two instances of potential rudder control problems in the accident airplane just days before the crash, both of which involved a component of the rudder control system called the **yaw dampers**. ([18], p.15)

The 737 yaw damper system is a type of autopilot that improves flight comfort by sensing and correcting minor yawing movements of the airplane due to turbulence. It contained two major parts: (1) a yaw damper coupler, which senses aircraft movements about the yaw axis, converts the motion to an electrical signal, and then sends it to the rudder main power control unit (PCU); (2) a transfer valve that receives the electrical signal from the yaw damper coupler, and converts it to mechanical inputs to the rudder main PCU, which in turn moves the rudder. ([9], p.17)

Despite their usefulness, yaw dampers sometimes malfunctioned and produced unnecessary rudder movements, and Boeing had never been able to resolve the problems successfully. Instead, Boeing recommended switching off the yaw damper whenever abnormal yawing happened. ([18], p.13) In addition, Boeing reduced the yaw damper's rudder movements to a maximum of 3 degrees to the left or right, so that the danger of rogue rudder movements due to yaw damper malfunctions could be minimized. ([18], p.14)

On February 25, 1991, six days before the crash, the accident airplane was at 10,000 feet when it yawed uncommanded as if the right rudder had been applied. The pilots turned off the yaw damper, and no further incident occurred during the flight. After this incident, United mechanics replaced the yaw damper coupler. However, on February 27, 1991, another crew on the accident airplane experienced uncommanded yawing to the right, and they also switched off the yaw damper to eliminate the problem. Afterward, the mechanics replaced the yaw damper transfer valve. Four days later, the Flight 585 accident happened. ([18], p.15)

The maximum authority of the yaw dampers was only 3 degrees of rudder movement (on each side). Consequently, a yaw damper malfunction might account for the uncommanded yawing motions in the two pre-crash incidents, but was unlikely to result in a loss of control of the airplane. ([12], p.87) Nevertheless, the discovery of these incidents was troubling for the following reasons: First, subsequent tests of the two removed components of the yaw damper showed that they function normally. ([12], p.87) Second, the yaw damper is connected with the rudder main power control unit (PCU), which can command a maximum of 26 degrees of rudder movements to the left or right. ([9], p.15) Could it be that these two pre-crash incidents were symptoms that reflected underlying issues of the main rudder PCU?

The second concerning discovery about the rudder control system came from the Systems Group. It involved another component in the rudder control system known as the **standby rudder power control unit**[13]. To understand the problem found by the Systems Group, it would be helpful first to understand the internal mechanism of a rudder power control unit (PCU).

In the Boeing 737-200, a rudder PCU controls the deflection of the rudder. It takes commands through cables from the cockpit and through electrical signals from the yaw damper, and directs the flow of pressurized hydraulic liquid in it to push the rudder left, right or neutral. The rudder PCU is about the size and shape of an upright vacuum cleaner, and consists of three main parts:

1. At the heart of the PCU is a soda-can shaped valve called the **dual concentric servo**

---

[13] In the NTSB report of Flight 585 accident, the standby rudder power control unit (PCU) is often called the standby rudder actuator. I avoided that terminology because I use the term "actuator" to refer specifically to the actuating cylinder (the tube with a piston to move the rudder), which is only a part of the standby rudder PCU.

valve. The servo valve is essentially a hydraulic switch and can trigger a powerful flow of hydraulic liquid in it to flow in two directions. One direction of the flow turns the rudder right, and the other direction turns the rudder left. ([18], p.119) The term "dual concentric" refers to the fact that there are two internal tubes (often called "slides") inside the servo valve housing: A **primary slide** moves within a **secondary slide**, which in turn moves within the servo valve housing. The design of the concentric slides aims for redundancy: If one slide jammed, the other could still move and carry out rudder commands. ([9], p.20.)

2. The inputs to the dual concentric servo valve come through a series of intricate mechanical linkages, including the so-called quadrant, torque tube, input rod, input crank, and input shaft. This series of linkages converts the rudder command input (either from the cockpit or from the yaw damper) into movements of two levers inside the servo valve: one lever (called the **primary summing lever**) moves the primary slide, and the other lever (called the **secondary summing lever**) moves the secondary slide. ([9], p.20)

3. A hydraulic cylinder called the **rudder actuator** receives the outputs of the dual concentric servo valve. The actuator consists of a hollow cylindrical tube along which a piston can slide; the piston connects to an actuator rod, which attaches to the rudder panel. Through holes on the cylinder wall, the dual concentric servo valve injects pressurized hydraulic fluid into one side of the actuator piston chamber and evacuates fluid from the other side. The pressure differential causes the piston to move in one direction, which moves the rudder via the connecting rod. ([18], p.120.)

To provide redundancy, two rudder PCUs are attached to the vertical fin structure of the 737-200. The main rudder PCU powered by two independent hydraulic systems assumes normal control of the rudder. A standby rudder PCU, normally unpressurized, provides backup just in case the main rudder PCU malfunctions. ([9], p.13) Each rudder PCU has its own servo valve, input linkages, and cylinder actuator.

Moreover, the input linkages (the input rods and cranks) to the two PCUs are connected through a mechanical device called the torque tube. As a result, in normal operations, the input linkages to the two rudder PCUs move in tandem in response to the rudder commands. The main rudder PCU then does all the work to move the rudder, while the standby rudder PCU components are simply moving along freely. Even though the standby rudder PCU does no work in normal operations, the free movements of its mechanical input linkages are important, because they play a part in the feedback loop that feeds the rudder movement back to the main rudder PCU. Because of the feedback loop, when the rudder surface deflects to the position commanded by the pilots, the input linkages to the main rudder PCU return to their neutral positions, so that the rudder does not move further than commanded. ([9], p.62)

When the Systems Group investigators examined the standby rudder PCU recovered from flight 585, they found evidence that a rotating component of the input linkage, called the **input shaft**, had jammed against the bearing through which the shaft is supposed to rotate.[14] The jamming of the two components was supported by the fact that **galling**, the transfer of metal from one component to the other, had occurred. ([9], p.50)

The discovering of the jamming raises concern for the following reason: As described earlier, the input linkages to the standby rudder PCU, the input linkages to the main rudder PCU, and the rudder itself constitute a complex feedback mechanism. When the main rudder PCU is in operation, the standby rudder PCU input linkages are supposed to move freely. If, however, the input linkages to the standby rudder PCU are not free to move due to jamming, that would change the geometric relationship between the input linkages to the main rudder PCU and the rudder. Could it be that the jamming of the input shaft

---

[14] The input shaft extends through, and rotates within, a bearing threaded within the body of the standby rudder PCU, and it is driven by another component in the input linkage called the input crank. In normal operations, the standby actuator does no work. However, both the input crank and the input shaft rotate freely to accommodate the relative motion between the rudder and torque tube. The end of the input shaft attaches to the summing levers of the servo valve. ([9], 50)

in the standby rudder PCU somehow locked up part of the complex feedback mechanism, and created a sufficiently large rogue movement in the main rudder PCU control system? ([9], p.62-63) The System Group investigators decided to examine this possibility.

While the Maintenance Group and the Systems Group were pursuing leads regarding the rudder control system, another group within the investigation—the Meteorology Group—zoomed in on another significant lead, namely severe atmospheric disturbances near the Colorado Springs Municipal Airport at the time of the accident.

Colorado Springs is located at an altitude of more than 6,000 feet above sea level. The terrain is flat and nearly featureless on three sides; to its west, however, are the mountains of the Front Range. ([18], p.40) When a rapid wind flows over the mountain ridge, sometimes it produces oscillations of the air called mountain waves in the downwind side of the ridge, causing moderate to severe turbulence. Moreover, often turbulent vortices of air, with a horizontal axis of rotation, form near the trough of the mountain on the downwind side. These turbulent horizontal vortices are called **rotors**, and they can cause unexpected changes in the strength and direction of the wind near the surface, sometimes in the opposite direction of the prevailing winds. Worse, rotors are only visible when there is sufficient moisture in the air. Otherwise, they offer no clear visual clue. ([18], p.46)

When the NTSB investigators examined the weather reports and witness reports of the day of the accident, they found indications of mountain waves and rotors. Hours before the accident, multiple pilots reported moderate to severe turbulence near Colorado Springs and Denver, and these reports were sent to nearby aircraft (including the accident aircraft) via in-flight weather advisory messages. ([9], p.27-28) Moreover, near the Colorado Springs Municipal Airport, arriving and departing aircraft experienced low-level wind shear in the form of sudden gains and losses of airspeed. ([9], p.23) Finally, multiple witnesses reported seeing rotor clouds or hearing roaring sounds characteristic of rotors on the day of the accident. ([9], p.35)

Nevertheless, even though mountain waves and rotors are hazards for aviators, it is uncommon for rotors of average strength to cause a complete loss of control in an aircraft. Is it possible that an unusually severe rotor brought down flight 585 near the airport? If so, is there any evidence of such a rotor on the day of the accident? These questions would occupy the Meteorology Group and the Performance Group in the following days of the investigation.

4.5 Examining the Rudder PCUs

Because the jamming of the standby rudder input shaft against its bearing appeared to be a possible explanation for the loss of control, the Systems Group investigators conducted a detailed examination of the input shaft and its bearing inside the standby rudder PCU to evaluate this scenario. The investigators wanted to know when the input shaft jammed to its bearing, how much binding force the jamming could produce, and whether pilot inputs to the rudder pedals could overcome the binding force.

First, the investigators addressed the question about the timing of the standby rudder input shaft-to-bearing jam. Since the galling (a type of wear caused by adhesion between sliding surfaces) between the input shaft and its bearing was the main evidence for the jamming, the investigators sought to determine whether galling occurred during the accident or before the accident.

During the examination, the investigators reassembled the input shaft and the bearing into the standby rudder PCU body. Based on the fire witness marks and soot patterns on the PCU body versus the input shaft and bearing, they determined the position of the bearing relative to the PCU body before the unit was exposed to the post-crash fire. They found that the bearing had been backed off (unscrewed) about 30 degrees of rotation from its fully seated position. At the backed off position, the bearing and the input shaft were free to rotate without any interference to the rest of the rudder control system. Moreover, the galled part of the bearing and the input shaft could be aligned only when the bearing was fully seated (without any backing off). ([9], p.64) These discoveries suggested that (1) galling occurred before the accident rather than within it and (2) at the end of the accident (and likely during the accident as well), the input shaft of the standby rudder PCU was free to rotate. ([9], p.122)

Second, the investigators conducted tests at Boeing's facility to determine the effects of a possible jam (consistent with the galling pattern found on the accident airplane) on aircraft controllability. Specifically, the goal of the tests was to estimate the binding force produced by the galling found on the accident airplane's standby rudder input shaft. For the tests, the experimenters custom manufactured several shaft-bearing pairs. To produce jamming, the clearance between the parts in each pair was much smaller than normal. Then for each pair, they tested how much binding force could be generated. After each test, they disassembled the parts and measured the galling pattern's surface area on each specimen using a microscope. Finally, the experimenters plotted a graph of the binding force versus the galled area recorded in the tests, and compared it to the measured area of gall in the accident shaft and bearing. ([9], p.65)

The test results showed that based on the areas of galling on the input shaft and bearing from the accident airplane, the galling could generate was about 70-80 pounds of binding forces at the end of the input crank, which is another component in the input linkages that drives the rotation of the input shaft. ([9], p.65) Translating this number to the pilot's pedal forces: pedal forces of only 35 pounds would be sufficient to overcome the binding force and regain control of the rudder. ([9], p.65) In short, the galling of the accident input shaft and bearing was unlikely to cause controllability problems.

Third, as a participant in the investigation, United Airlines also inspected other 737s to see if there were other examples of standby rudder input shaft-to-bearing galling. It found one 737-200 with a galled input shaft and bearing, and its bearing was also backed off (unscrewed) for about 20 degrees. The NTSB metallurgists characterized the galling in this airplane as worse than the one on the accident airplane, and yet there was no evidence that the galled components in this aircraft had ever caused detectable problems. ([9], p.122)

Finally, to determine if the input shaft jamming scenario could explain the two incidents of uncommanded rudder movements experienced by the accident airplane before the crash, the investigators examined the combined effects of a yaw damper rudder input and the binding of the standby rudder input shaft to its bearing. The results of the examination showed that such a combination could produce an uncommanded rudder deflection of at most 5.5 degrees. Further simulation tests showed that the pilot's control inputs could easily counter 5.5 degrees of rudder movement. ([9], p.121) In short, the input shaft jamming scenario is consistent with the two rudder-related incidents prior to the crash, but it is insufficient to cause the airplane to lose control.

Based on these considerations, the Systems Group investigators concluded that the jamming and galling between the standby rudder input shaft and its bearing was not a contributory factor to the loss of control of the accident airplane. ([9], p.122)

In addition to the standby rudder PCU, The Systems Group also examined the main rudder PCU from Flight 585 in search of pre-crash anomalies that could have contributed to the accident. However, the main rudder PCU was substantially damaged by impact and post-crash fire, making it impossible to test it as a complete unit. ([9], p.49) As a result, the investigators disassembled the dual-concentric servo valve of the main rudder PCU. They brought the components of the servo valve to a facility of Parker Hannifin—the manufacturer of the main rudder PCU—in Irvine, California for testing. ([35])

At the Irvine facility, the investigators discovered that three parts of the dual concentric servo valve were missing: A spring, a spring guide, and an end cap. Together, these three components served as an internal stop to the two moving slides within the servo valve housing. They were supposed to be packed alongside the other servo valve components, and the investigators could not identify why they were missing. ([35])

The Systems Group decided to test the servo valve using a new spring, spring guide, and end cap from Parker Hannifin's storage. After polishing the servo valve's interior housing wall and the surfaces of the two internal slides, the investigators placed the slides back inside the servo valve housing, along with the new spring, spring guide, and end cap. When they tested this restored servo valve, they found that it failed to maintain pressure according to specifications. ([35]) However, the investigators concluded that the failure to pressurize was likely due to crash impact because the flight crew would have detected it if it had occurred inflight. ([9], p.49) The tests did not reveal other anomalies that could have contributed to the accident.

4.6 Reconstructing the Weather

During the initial investigation, the NTSB Meteorology Group carried out the routine post-accident task of reconstructing the general weather condition on the day of the accident by conducting witness interviews and checking local weather records, forecasts, and weather warnings. ([18], p.48) The group was particularly interested in the wind conditions that day, as they could be relevant to the accident.

The investigators found conflicting wind speed data in the weather records and forecasts. They found that the wind speeds transmitted by the control tower broadcasts were significantly stronger than those recorded on automatic devices. Before the accident, the control towers at the Colorado Springs municipal airport broadcast wind speed about 20-22 knots, with gusts to 30 knots. In contrast, the six sensors of the Low-Level Windshear Alert System (LLWAS) located at the airport recorded only winds of 13 knots, with gusts to 23 knots, minutes before the accident. The discrepancy between the broadcast and the recorded wind data was puzzling because the local controller stated that the LLWAS was the only instrument the control tower relied on for wind speed. In the end, the investigators could not determine the reason for the discrepancy between the broadcast and recorded wind speed data. ([9], p.23-24)

The witness reports were more suggestive and pointed to the existence of windstorms[15] and rotors[16]. Even though most witnesses near the crash site only reported light winds at the time, several witnesses within a few miles of the crash site reported short gusts with estimated speed ranging from 50 miles to 90 miles per hour. These gusts could be the result of rotors hitting the ground. ([9], p.34) Moreover, some witnesses observed a few rotor clouds within 15 miles of the Colorado Springs airport. One person about 6 miles from the crash site reporting seeing rotor clouds 10-15 minutes before the crash, although he was unsure of their intensity. ([9], p.35) All in all, the witness observations were not precise enough to settle the question about whether rotors and other mountain-induced wind events were present at the time and location of the accident. However, they did suggest the need to pursue this question further.

Beginning March 27 (24 days after the accident), the NTSB's Meteorology Group convened a series of meetings of outside meteorological experts. The outsiders present included scientists from the National Oceanic and Atmospheric Agency (NOAA), the National Center for Atmospheric Research (NCAR), and the University of Wyoming Atmospheric Science Department. One of the meetings' goals was to examine what mountain-generated violent wind events could have occurred on the day of the accident. ([9], p.32)

The experts presented various scenarios in which rapid air-streams blowing over a mountain ridge generated severe windstorms and rotors on the downwind side. They estimated the typical strength of a rotor on the day of the accident, and agreed that a representative rotor would have a radius of about 500 meters; a tangential velocity of 30 meters per second, and a rotational rate of 0.06 radians per second (or 3.5 degrees per second). ([9], p.32) Rotors were of particular interest because of how Flight 585 rolled to the right before the crash. However, the experts also considered the potential influence of other atmospheric phenomena such as "jumps", which were nearly vertical downdrafts of cold air that rebound

---

[15] A windstorm is a strong wind that causes at least some damage to trees and buildings.

[16] A rotor is a wind vortex that has a horizontal axis of rotation.

as strong updrafts. ([18], p.47)

The theories presented at the meteorological meetings did not form a coherent picture. Consequently, the NTSB commissioned the NCAR (National Center for Atmospheric Research) to use existing weather data to build a simulation model of the March 3 weather on the downwind side of the Front Range at Colorado Springs. The goal of the simulation was to locate possible severe wind events and see whether flight 585 could have encountered some of these events. ([18], p.48)

The weather simulation turned out to be more challenging than expected, partly because standard modeling assumptions about air flows over the mountains were not appropriate in this study due to some complex geographical features of the Front Range. ([9], p.53) In the end, the simulation study concluded that it was impossible to determine from modeling whether any particular transient wind events occurred on the day of the accident in Colorado Springs. Even at high resolutions, models could only suggest the general structure of windstorms, but were not precise enough to determine whether storms with specific features occurred at a particular location and time. At best, modeling could determine whether such severe wind events were possible. Based on the modeling results and observations of the weather in the area, the NCAR study concluded that severe wind events were *possible* on March 3, 1991, in Colorado Springs. ([9], p.54)

4.7 Simulating Flight 585

While the extensive set of Colorado Springs weather data turned out to be inconclusive, the NTSB's Performance Group turned to another important investigative tool, namely the flight simulator. Boeing has a multi-purpose cab (M-Cab) flight simulator is capable of simulating events outside of normal flight regimes. It has a faithful mock-up of the aircraft cockpit and is controlled by extremely sophisticated computer software to simulate a variety of flight scenarios. ([18], p.51) Between May 10, 1991, and April 28, 1992, the

Performance Group and other participating parties gathered at Boeing's M-Cab facility on four different occasions to examine the effects of atmospheric disturbances and flight control system malfunctions on the flight path of a 737-200 airplane. ([9], p.57)

For these simulation studies, Boeing developed simulator software for various scenarios of atmospheric disturbances or flight control malfunctions. A pilot would then attempt to maintain control of the simulated airplane while countering the atmospheric disturbance or control malfunction. Moreover, the pilot would attempt to follow the flight path of the accident airplane, as determined by FDR and radar data. The main goal of the flight simulations was to identify factors sufficient to cause an airplane to lose control and crash, in a way consistent with the flight path of Flight 585. ([9], p.57)

First, among scenarios of atmospheric disturbances, the main scenario of interest was an encounter with a rotor. To find out how strong a rotor needs to be to cause a loss of control of the aircraft, the investigators conducted a sequence of simulations in which the severity of the rotor was increased, until the pilot had "extreme control difficulties". ([9], p.58) They found that rotors that generated extreme control difficulties had rotation rates of 0.6 radians per second (34 degrees per second), a 250 feet core radius, and 150 feet per second tangential velocity. When the rotation rate decreased to 0.4 radians per second, and the tangential velocity decreased to 100 feet per second, the pilot had trouble controlling the airplane precisely. However, these control difficulties did not necessarily result in a crash. ([9], p.58)

Since experts from the meteorological meetings had established that a typical rotor on the day of the accident had a rotation rate of 0.06 radians per second, the investigators also tested the scenario of a typical rotor encounter. The simulations showed that the 0.06 radians per second rotors had little or no effect on the aircraft controllability, except for some loss of altitude and airspeed. ([9], p.57) Flight 585 did lose some altitude during its descent, but its airspeed remained mostly constant, until the final few seconds before the crash when the airspeed increased drastically. ([9], p.58) In short, rotors with representative speeds could not have caused Flight 585 to lose control. If a rotor brought down Flight 585, it must have been unusually severe, with a rotation rate ten times as fast as the typical rotors.

Moreover, in addition to being unlikely, the theory that Flight 585 encountered an unusually severe rotor faces an additional problem. Similar to tornadoes, rotors usually have an area of low atmospheric pressure at their core. The faster a rotor rotates, the more dramatic the pressure drop inside it is. Moreover, an airplane's altimeter that records altitude is essentially a barometer that converts atmospheric pressure into altitude values. As a result, when an airplane penetrated the low-pressure core of a rotor, its altimeter could be fooled into recording a sudden increase in altitude. ([18], p.52) According to NTSB's estimation, if Flight 585 had encountered a rotor strong enough to cause airplane control problems, a transient increase in altitude of several hundred feet should have been recorded on the FDR. However, the FDR data showed no such altitude spikes, but only a steady curve of slow descent followed by a sudden plunge before the crash. ([9], p.127)

In addition to rotors, Boeing simulated a few other atmospheric phenomena that could have occurred at the time of the crash. One such phenomenon is the so-called jumps—concentrated streams of wind with upward vertical motion. The simulation showed that jumps could not have contributed to the accident because their wind shear values (i.e., their abilities to cause rapid changes in wind speed and direction over short distances) were too low. ([9], p.127)

Second, among scenarios of flight control malfunctions, the main scenario of interest was the occurrence of a **rudder hardover**. A rudder hardover is a scenario in which the rudder deflects to its full travel positions to the left or right at a given flight condition, and the maximum angle of rudder deflection (from neutral to full left or right) is called the **blowdown limit**.[17] As more and more alternative scenarios of flight control malfunctions

---

[17] Rudder blowdown limit represents a balance between the aerodynamic forces acting on the rudder and the mechanical forces produced by the PCU. On the ground, a 737 main rudder PCU can command a maximum deflection of 26 degrees off the rudder's neutral position. When in flight, the blowdown limit of

were shown to be insufficient to cause control problems, a rudder hardover to the right became one of the few remaining possibilities that might explain Flight 585's sudden right roll and yaw before the crash. ([18], 52-53) So the NTSB requested Boeing to conduct flight simulations concerning the effects of rudder hardover on flight controllability. ([9], p.60)

According to the Boeing simulations, moderate to high rudder deflection rates almost always result in large and rapid heading changes. Moderate uncommanded rudder deflections (e.g., deflection to 7.5 degrees) could be easily countered with pilot control wheel inputs (to the ailerons and spoilers). The controllability of uncommanded rudder hardover, however, was more difficult to determine. The simulations suggested that depending on the configurations of other control surfaces, sometimes a rudder hardover could result in loss of control and ground impact even if full control wheel countermeasures by the pilots were applied. Even for cases where a crash could be avoided, immediate, full control wheel deflections were required to prevent the crash. ([9], p.60)

However, the rudder hardover scenario raised a further question that was difficult to answer. The flight simulations showed that *if* a rudder hardover had occurred in Flight 585, it *could* have been sufficient to cause the aircraft to lose control and crash. However, what could have triggered a rudder hardover in Flight 585? Based on the FDR data of Flight 585, the most likely time for a rudder hardover to occur was at about 09:43:30 (11 seconds before the crash), when a rapid heading change to the right at a rate of 4.7 degrees per second occurred. Based on the CVR data, it was highly unlikely that the pilots triggered such a rudder hardover by pressing down the right rudder pedal until the crash.

Moreover, the NTSB could not identify a failure mechanism within the rudder control system that could produce an uncommanded rudder hardover: First, the yaw damper by itself could trigger at most 3 degrees of rudder deflection, a far cry from a rudder hardover (up to 26 degrees). Similarly, the galling found in the standby rudder PCU was not sufficient to cause a full rudder deflection. Finally, the main rudder PCU was too damaged to be

---

the rudder is reduced.

tested as a complete unit, and the tests on its components so far had failed to reveal precrash anomalies that could have caused control problems. ([9], p.49)

In short, results from the flight simulations were again inconclusive. Both the encounter with an extremely severe rotor and the occurrence of a rudder hardover could have caused Flight 585 to lose control and crash. However, neither hypothetical event seemed likely given what the investigators knew at the time.

4.8 The Mack Moore Incident

On July 16, 1992 (about a year and four months after the Flight 585 crash), during a preflight rudder control ground check at Chicago O'Hare International Airport, a captain (named Mack Moore) of a United Airlines 737-300 made a surprising discovery. When he moved the left rudder pedal more rapidly than usual, he discovered that it stopped and jammed near a quarter of the way down. When he released pressure, the pedal returned to its neutral position. Moore returned the airplane to the gate, and the United mechanics removed the main rudder PCU of the airplane for further examination. ([9], p.69)

The main rudder PCU of the incident airplane was first tested at United Airlines' facility and then at Parker (the manufacture)'s facility. The tests found multiple anomalous behaviors, ranging from internal leakage of hydraulic fluid to the sluggish movements of the rudder actuator's piston. However, the most alarming discovery was that this PCU could generate a type of rudder movement called **rudder reversal**, which means that the rudder would move in a direction opposite to the one commanded by the pilot. ([9], p.69)

The main rudder PCU is a hydraulic switch. It takes inputs from the rudder pedals and the yaw damper, and translates them into hydraulic flows that turn the rudder. Inside the PCU, the device called dual concentric servo valve directs hydraulic flow through a set of tiny channels or "ports" on its housing wall to an actuating cylinder. The hydraulic flow moves a piston inside the actuating cylinder, which turns the rudder left or right. Inside the housing of the dual concentric servo valve, there are two concentric sliding tubes called the primary and the secondary slides, with the primary slide moving inside the secondary slide. The perimeters of both slides have a set of passageways on them; depending on the rudder command inputs, the passageways can line up in different ways with the channels on the servo valve housing wall. Lined up in one way, the hydraulic flow moves in one direction that turns the rudder right. Lined up in another way, the hydraulic flow moves in a different direction that turns the rudder left. ([18], p.57)

When the investigators examined the dual concentric servo valve of the main rudder PCU from the Mack Moore incident airplane, they discovered that the secondary slide could overtravel beyond its designed operating position. The result of the overtravel was that the passageways on the primary and secondary slides lined up with the wrong ports on the servo valve housing, causing an abnormal hydraulic flow in the opposite direction of the intended one. ([9], p.69.)

However, what enabled the secondary slide from the incident airplane to overtravel? In subsequent tests, the investigators identified a few conditions necessary for the secondary slide from the incident airplane to overtravel. First, the rudder pedals were pressed rapidly to command a maximum rate of rudder deflection. Second, due to a manufacturing defect, the secondary summing lever (the lever inside the servo valve that moves the secondary slide) could not maintain contact with its external stop. Given both of these conditions, the secondary summing lever would move beyond its external stop, which allowed the secondary slide to travel beyond its design limits. ([9], p.69)

The next question then became: Did a rudder reversal occur in Flight 585? To help answer this question, the NTSB Systems investigators subjected the PCU servo valve from Flight 585 to further tests to see whether abnormal movements of the concentric primary and secondary slides in this PCU were possible. The results were disappointing: Unlike the servo valve from the Mack Moore incident, the secondary summing lever from the Flight 585 servo valve was within the design specification and made full contact with its external stop. As a result, the secondary slide from the servo valve of Flight 585 did not overtravel in the tests. The investigators could not generate a rudder reversal via the mechanism found in the Mack Moore incident. ([9], p.70)

4.9 The Lack of a Conclusion

On December 8, 1992, 21 months after the crash, the NTSB officially declared at a public board meeting that Flight 585 had crashed for "undetermined reasons". ([18], p.66) The "probable cause" statement of the accident report adopted on that day reads as follows:

> The National Transportation Safety Board, after an exhaustive investigation effort, could not identify conclusive evidence to explain the loss of United Airlines flight 585.
>
> The two most likely events that could have resulted in a sudden uncontrollable lateral upset are a malfunction of the airplane's lateral or directional control system or an encounter with an unusually severe atmospheric disturbance. Although anomalies were identified in the airplane's rudder control system, none would have produced a rudder movement that could not have been easily countered by the airplane's lateral controls. The most likely atmospheric disturbance to produce an uncontrollable rolling moment was a rotor (a horizontal axis vortex) produced by a combination of high winds aloft and the mountain terrain. Conditions were conducive to the formation of a rotor, and some witness observations support the existence of a rotor at or near the time and place of the accident. However, too little is known about the characteristic of such rotors to conclude decisively whether they were a factor in this accident. ([12], p.102)

For only the fourth time in its history, the NTSB had failed to determine the probable cause of a plane crash. ([18], p.67)

Even though it could not determine why Flight 585 crashed, the NTSB was concerned about what it had found about the Boeing 737 rudder problems. During the investigation of Flight 585 crash, the NTSB issued a total of five safety recommendations to the FAA (Federal Aviation Administration) regarding the rudder control system. ([18], p.67)

The first of the rudder-related safety recommendation was issued on August 20, 1991, due to NTSB's concern about galling found on the standby rudder PCU input shaft and its bearing. The safety recommendation asks the FAA to issue an Airworthiness Directive (AD)[18], which requires the airlines to check the force needed to rotate the input shaft in the standby rudder PCU relative to its bearing. If excessive force was needed to rotate the input shaft or the bearing itself also rotated along with the input shaft, the parts should be immediately removed from service and replaced with parts that have more clearance between them. ([12], p.104)

The other four rudder-related safety recommendations were issued on November 10, 1992, in the wake of the Mack Moore incident. The recommendations asked the FAA to require Boeing to design a maintenance test procedure to detect anomalous movements inside the main rudder PCU servo valve. In the meantime, the recommendations suggested that Parker Hannifin be required to redesign the main rudder PCU servo valve, so that the possibility of overtravel of the secondary slide could be eliminated. The airlines should also be required to replace the main rudder PCUs in their 737 fleets as soon as the new designs were available. ([12], p.103)

After a review of the 737 rudder system, the FAA decided that the galling and jamming of the standby rudder PCU input shaft could easily be detected by the pilots and did not warrant issuing an AD. ([18], p.69) The FAA did, however, issued two ADs in 1994, requiring a main rudder PCU inspection every 750 hours of flight time until the new designs became available for replacement. The inspection procedure was specifically designed by Boeing to identify internal leakage within the main rudder PCU servo valve, which is a symptom of the secondary slide overtravel. ([18], p.70)

By August 14, 1994, the NTSB had classified FAA's responses as acceptable and had closed the file on all the five rudder-related safety recommendations. Less than a month later, the USAir Flight 427 accident happened. ([18], p.71)

---

[18] An Airworthiness Directive (AD) is a notice to the manufacturer and operators of a certified aircraft that a known safety deficiency exists with the aircraft and must be corrected. Compliance with the AD is mandatory and legally enforceable by federal law.

Chapter 5

Case Study 4: USAir Flight 427

5.1 History of the Flight

On September 8, 1994, USAir[1] Flight 427, a Boeing 737-300 aircraft[2], crashed near Pittsburgh International Airport (PIT), Pittsburgh, Pennsylvania. The impact destroyed the aircraft and killed all 132 people on board. ([10], p.1)

The accident flight was on the last day of a 3-day trip for the flight crew. At about 18:02, September 8, Flight 427 departed the gate at Chicago O'Hare International Airport and became airborne at about 18:10. The destination was Pittsburgh International Airport, and the estimated en route time was 55 minutes. The first officer was the pilot flying, whereas the captain handled radar communications and other pilot-not-flying (PNF) tasks. ([10], p.1)

Shortly before 18:57, the captain contacted Pittsburgh's control tower and stated that they were descending to 10,000 feet above sea level. The Pittsburgh approach controller responded at 18:57:23, stating that Flight 427 would receive radar vectors to the final

---

[1] USAir was a major airline in the U.S. In early 1997, USAir changed its name to USAirways. In 2013, USAirways merged with American Airlines.

[2] The Boeing 737 evolved through generations. The Models 737-100 and 737-200 belong to the first generation, whereas the 737-300 model belongs to the second generation.

approach for runway 28 Right (28R) at PIT. About 18:58:03, the controller instructed Flight 427 to descend and maintain an altitude of 6,000 feet. ([10], p.2-3) According to ATC (Air Traffic Control) and radar information, the control tower had also been in contact with Delta Airlines Flight 1083, a Boeing 727 that preceded USAir flight 427 on the approach to PIT from the northwest. ([10], p.2)

About 19:02:22, the tower told the USAir flight 427 crew to turn left and adopt a heading of 100 degrees. In addition, the controller stated that there was nearby traffic, a northbound Jetstream aircraft climbing to 5,000 feet about 6 miles away. About 19:02:32, the captain acknowledged the instructions from the controller, stating that they were looking out for the traffic and turning to 100 degrees of heading. ([10], p.3) That was the last routine transmission from flight 427.

According to the Air Traffic Control (ATC) tapes, at about 19:03:10, a radio transmission from USAir flight 427 stated: “Oh (unintelligible) Oh (unintelligible).” The controller reported that at the time, flight 427’s altitude readout on the radar screen was 5,300 feet. About 19:03:14, the controller advised, “USAir 427 maintain 6,000, over.” About 19:03:15, the captain of flight 427 made a radio transmission, stating “four twenty seven emergency!” From about 19:03:09 to 19:03:22, the first officer’s radio microphone was activated and deactivated repeatedly, and the ATC tapes recorded exclamations of the pilots and other sounds from flight 427. ([10], p.6)

Meanwhile, it was still daylight hours in Pittsburgh. Many witnesses saw an aircraft appearing to turn left at a low altitude, but its right wing continued to lift until it was perpendicular to the ground. Then the aircraft turned upside down and dived towards the ground. ([18], p.76)

At 19:03:23 eastern daylight time, USAir flight 427 impacted a hilly terrain near Aliquippa, Pennsylvania, about 6 miles northwest of PIT. The air traffic controllers in the control tower cab reported seeing heavy smoke rising to the northwest of the airport. ([10], p.6)

5.2 Initial Investigation

The NTSB's Go Team arrived at Pittsburgh on the night of the accident and held its first official meeting the next day. ([18], p.77-79) The meeting established the leaderships in each of the investigative groups such as Structures, Powerplants, Systems, and Operations. It assigned representatives from various outside parties such as FAA[3], USAir, Boeing, ALPA[4] and Parker Hannifin[5] to appropriate groups based on their expertise. After a walkabout at the crash site, each investigative group began to carry out its assigned tasks. ([18], p.80-81)

Three groups managed the crash site in addition to documenting and recovering different parts of the wreckage. The Structures Group focused on the fuselage and airframe; the Powerplants Group focused on the engines and engine accessories; the Systems Groups focused on the electrical and hydraulic system's components. ([18], p.79) Because the coroner's staff had to disturb the wreckage to find human remains, investigators from the three groups had to work alongside them and document the wreckage distribution before it was disturbed. ([18], p.84) The on-site phase of the investigation occurred between September 9 (one day after the crash) and September 20, 1994. ([10], p.36)

Flight 427's impact site was in a densely wooded area on an upward sloping hill. The wreckage was severely crushed and fragmented, and most of it was in the main impact crater with a 350-foot radius. ([10], p.36) Some parts of the wreckage were buried so deeply that they were not visible above the ground. The investigators had to use metal detectors and ground-penetrating radar equipment loaned by the U.S. Bureau of Mines to find some pieces. They examined and documented each piece of the recovered wreckage at the crash site, looking for any signs of pre-impact damage. After the wreckage was documented and

---

[3] FAA stands for Federal Aviation Administration. It is a U.S government body that regulates all aspects of civil aviation in the U.S. After each aviation accident investigation, the NTSB issues safety recommendations to the FAA, and the latter determines how to act on the recommendations and whether to turn them into regulations.

[4] ALPA stands for Air Line Pilots Association. It is a union representing pilots from over 35 U.S and Canadian airlines.

[5] The Parker-Hannifin Corporation, often abbreviated to as Parker, is the manufacturer of many of Boeing 737's hydraulic system components.

decontaminated, it was relocated from the accident site to a hanger at PIT for further examination. The exceptions were the black boxes, which were sent to NTSB's labs in Washington D.C. for readout. ([10], p.38)

In addition to the main impact crater, the on-site investigators carefully examined the surrounding landscape for clues about how the plane impacted the ground. The left wing rested midway between the main wreckage and a 25-foot long scar on the ground, suggesting that it had been the first part of the airplane to hit the ground. Vertical breaks of branches on a tree beside the scar indicated that the wing was almost vertical when it struck the tree. The trail of tree damage before the scar provided further information about the direction and angle of the ground impact. The examinations showed clearly that the airplane had crashed at a steep angle. ([18], p.85)

Finally, to determine whether anything had separated from the accident airplane before its crash, the on-site investigators conducted a ground and helicopter search in a large ground area over which the aircraft had passed shortly before the crash. They did not find any significant airplane components away from the main crash site, although some lightweight pieces such as insulation and paper were located as far as 2.5 miles downwind from the crash site. All the four corners of the airplane[6] and all the flight control components were at the main wreckage site. All but one out of over 100 witnesses stated that the airplane appeared to be intact before the crash sequence. Consequently, the investigators concluded that the lightweight pieces likely became airborne during the crash explosion and drifted downwind afterward. ([10], p.240)

While the Powerplants, Structures, and Systems Groups were securing the crash site and examining the wreckage, the Operations Group turned its attention to the background information concerning the flight crew, the accident airplane, and the airline procedures and practices at USAir. For instance, the Operations Group wanted to know: ([18], p.87; [10], p.240)

---

[6] The four corners of an airplane are its nose, wingtips, and tail.

- Whether the flight crew were adequately trained and certified, and whether they had any pre-existing medical or behavioral conditions that might have contributed to the accident;
- Whether the accident airplane was equipped, maintained and operated following Federal regulations and industry-approved practices;
- Whether the maintenance records of the accident aircraft showed any recurring or un-rectified problems that might be relevant to the accident;
- How USAir 427 was flown on its final flight, what its detailed flight path was, and how the PIT control tower handled the flight;
- How the USAir pilots were recruited, how their performance was evaluated, how they were trained to handle emergencies, and what the airline's safety culture was.

Due to its broad scope, the Operations Group was divided into subgroups, each focusing on a subset of background questions. For instance, the Human Performance Group focused on the background and the performance of the pilots, the Maintenance Group examined the maintenance records, and the Cockpit Voice Recorder Group along with the Flight Data Recorder Group reconstructed details of the accident flight from tapes of the black boxes.

5.3 The Black Boxes

On the morning after the crash, the two black boxes—the cockpit voice recorder (CVR) and the flight data recorder (FDR)—from Flight 427 arrived at the NTSB's laboratories in Washington D.C. The two recorders survived the crash partly because they were located in the airplane's tail, which suffered the least damage. They were also extraordinarily strong, able to withstand an impact force of 2,400 Gs and a fire of 2,000 degrees Fahrenheit for thirty minutes. ([1], p.38)

The cockpit voice recorder (CVR) from the USAir plane recorded four channels of sounds: one from the cockpit area microphone (CAM) in the cockpit ceiling, one from each of the pilot's headsets, and one from an oxygen mask in the jump seat. ([1], p.38) Although the CVR unit suffered from external and internal structure damage, the magnetic recording tape was in good condition. The quality of the recording was excellent. Both pilots had worn "hot" mikes—headsets similar to the ones astronauts wear. The mikes were very close to the pilot's mouths so that they picked up every word. They even recorded the pilot's breathing. ([1], p.39)

The flight data recorder (FDR) also exhibited structural damage, although its solid-state flash memory module was intact and yielded good data. The FDR recorded a total of 13 flight parameters. Eight of the parameters were recorded at once-per-second intervals: altitude, airspeed, heading, microphone keying, and four parameters related to the engine. The other five parameters were sampled more frequently, between 2-8 times per second. These parameters included roll attitude[7], pitch attitude[8], control column[9] position, longitudinal acceleration[10], and vertical acceleration. The FDR did not record the positions of any flight control surfaces; in particular, it did not record what the rudder was doing or whether the pilots were pushing on the rudder pedals. ([10], p.35-36)

The microphone keying parameter recorded in FDR showed when the pilots were pushing a button to talk with the air traffic controllers. This parameter allowed the investigators to synchronize the FDR data with the CVR data and air traffic control (ATC) data. ([1], p.39) After correlating these three sources of data, the investigators obtained the following picture of the last minutes of Flight 427:

The accident flight was routine until it was near PIT. At 19:02:53, about 30 seconds before the crash, USAir flight 427 was rolling out of a 7 degrees left bank towards a wing-level

---

[7] Roll attitude is the degree to which the wings were rolling left or right, relative to the horizontal plane.

[8] Pitch attitude is the degree to which the nose is pointing up or down, relative to the horizontal plane.

[9] By turning the control column forward or aft, the pilots control the pitch attitude of the airplane.

[10] Longitudinal acceleration is the acceleration along the longitudinal axis, i.e., the nose-to-tail axis.

attitude. It approached the ATC-assigned heading (100 degrees), airspeed (190 knots), and altitude (6,000 feet msl[11]). At 09:02:54, the first officer said, “Oh ya, I see zuh Jetstream[12].” ([10], p.4)

As the first officer finished this statement (at about 19:02:57), the CVR recorded the sound of three thumps in 1 second, the captain stating “sheeez”, and the first officer stating “zuh”. Over the next three seconds, the aircraft’s left bank steepened from about 8 degrees to about 20 degrees, then decreased slightly to about 15 degrees, but never reached a wing level attitude. During these three seconds, the CVR recorded two more thumps, two “clickety-click” sounds, the sound of the engine noise getting louder, and the sound of the captain inhaling and exhaling quickly. ([10], p.4)

About 19:02:59 (about 23 seconds before the crash), the airplane’s heading, which had been moving left steadily towards the ATC-assigned 100 degrees, began to move left at a faster rate and passed through 100 degrees. At this time, the CVR recorded the captain stating “whoa” and the first officer grunting softly. ([10], p.4)

By 19:03:01, the airplane’s heading had moved past 89 degrees and continued to move left at least 5 degrees per second. The airplane had begun to roll rapidly back to the left again, and just before 19:03:03, the left bank angle had increased to about 43 degrees. The aircraft’s pitch attitude began to decrease rapidly. The airspeed started to decrease below the assigned speed of 190 knots, and the FDR recorded an increased amount of aft control column command while the autopilot maintained level flight. Between about 19:03:01 and about 19:03:04, the CVR recorded the sound of the first officer grunting loudly and making brief exclamations, while the captain repeatedly stating “hang on”. ([10], p.4)

At 19:03:04, the CVR recorded the sound of the autopilot disconnect horn. ([10], p.4) At that moment, the rate of descent was about 2,400 feet per minute. The left yaw and roll continued to increase. By 19:03:05, the aircraft reached a left bank of 55 degrees. Its nose

---

[11] MSL stands for mean sea level, it represents the baseline elevation relative to which altitude is measured.

[12] The “jetstream” refers to a nearby Jetstream aircraft that the ATC mentioned to the pilots. The first officer made this statement in an accent, presumably as a joke.

was 10 degrees below the horizon, and its rate of descent was about 3,000 feet per minute. The control column was moving aft, and vertical acceleration increased. ([5])

At 19:03:07 (15 seconds before the crash), the airplane's nose was nearly 20 degrees below the horizon. The left bank angle had increased to 70 degrees, and the descent rate was approximately 3,600 feet per minute. ([5]) Meanwhile, the CVR recorded a sound increasing in loudness and similar to the onset of the stall[13] buffet. The captain asked, "what the hell is this?" Less than 1 second later, a vibrating sound similar to the aircraft stick shaker[14] started and continued until the end of the recording. The sound of the stick shaker was followed immediately by more warning sounds, including a sound similar to an altitude alert, and the sound of the traffic alert and collision avoidance system[15]. ([10], p.6)

The left yaw and roll continued, and the aircraft reached 90 degrees nose down at 19:03:13, approximately 3,600 feet above the ground. The aircraft continued to roll after reaching a vertical dive, but the nose began to rise. By about 19:03:16, the aircraft had completed a full 360 degree left roll. ([18], p.125) At this point, its nose was about 40 degrees down, and the left roll stopped briefly. At 19:03:18 (4 seconds before the crash), the Captain said, "pull...pull...pull", but the aircraft immediately resumed its left roll, and the nose dropped again. At 19:03:22, the aircraft impacted the ground at about 80 degrees of vertical dive, almost 60 degrees of left bank, and at 261 knots indicated airspeed. ([5])

The accident sequence extracted from the black boxes raised further questions. First, FDR data indicated that the accident sequence began at about 19:02:58 when Flight 427 began to yaw and roll at an increasing rate to the left. By 19:03:01, the airplane's heading was moving left at least 5 degrees per second until the stick shaker activated at 19:03:08. The airplane's left roll angle was also increasing rapidly during this time: Its left roll angle was about 28 degrees at 19:03:01 and exceeded 70 degrees five seconds later. What could

---

[13] A stall is a condition in which the wings can longer produce enough lift for normal flight.

[14] The stick shaker is a device that caused the control column to rattle as a warning when the aircraft is in danger of stalling.

[15] The traffic alert and collision avoidance system is a system based on radar beacon signals from aircraft.

have initiated and sustained such a drastic yawing and rolling movement?

To answer this question, the investigators had to examine many possible scenarios. The primary suspect, however, was the rudder, which controls rotation around the vertical or yawing axis of the aircraft. Unlike other two-engine large transport aircraft in service at the time, the Boeing 737 has a single, unusually large rudder panel. The single rudder design allows the deflection of the rudder to generate a powerful force, which could be strong enough to counter the yawing produced by the failure of a wing-mounted engine. ([18], p.119) However, it also means that an unintended rudder deflection could be sufficient to cause an aircraft to lose control. Moreover, the final flight profile of USAir Flight 427 is very similar to that of United Flight 585. Even though the Flight 585 case was unsolved at the time, the potential rudder failure mechanisms found in that investigation were still fresh in the memories of the NTSB investigators. ([18], p.127)

Given the rudder's potential importance in this accident, it would have been helpful if the flight data recorder (FDR) had recorded the rudder's position or the pilot inputs to the rudder pedals. Unfortunately, just like the FDR on Flight 585, the FDR on Flight 427 did not record any flight control surfaces' positions. Moreover, apart from the control column position, the FDR did not record anything about pilot inputs to flight controls. The data for the control column position only showed that the pilots were pulling back the control column up in an attempt to bring the aircraft's nose up, which did not help to answer the question about what the rudder was doing. ([18], p.125) In short, the investigators had to reconstruct the rudder positions and the pilot inputs to rudder pedals by other means.

Second, when the accident sequence started (about 19:02:57), the CVR recorded three mysterious thumping sounds in 1 second. The first few days after the accident, members of the CVR team listened to these sounds hundreds of times but could not recognize them. ([1], p.48) If the thumps had been heard at a different time on the CVR tape, the investigators probably would not have worried about them. After all, airplanes made so many noises that CVR tapes often had sounds that could not be identified. However, the investigators thought it was important to answer the question, “What was the source of the thumps?”, because the thumps occurred at a crucial time, right before the airplane started to roll left. ([1], p.136)

On September 11, three days after the crash, members of the CVR team visited a USAir plane at Washington National Airport. Their goal was to record various sounds on the plane’s CVR to see if they matched the thumps from Flight 427. The investigators recorded the sounds of various flight control manipulations; they dropped objects on the floor, stomped their feet on the doorway, and even triggered the stick shaker. However, when they ran the newly recorded tape through the computer, the acoustic fingerprints of the recorded sounds did not match the thumps. ([1], p.48)

The thumps were not the only mysterious sounds recorded at the beginning of the accident sequence. Immediately following the three thumps, the CVR recorded two “clickety-click” sounds, and the investigators were also unable to identify the sources of these clicks. Moreover, at both 19:02:58 and 19:03:02, the CVR recorded the engines increasing in volume. When the CVR team ran the engine sounds through a computerized waveform analysis, they found that the sound’s pitch remained unchanged, but its loudness increased. ([10], p.132) However, according to the engine parameters recorded on the FDR, the engines were running normally and were not throttled up when the accident sequence started. If so, why did the engine sounds get louder in the CVR? The CVR specialists suspected that the answer had something to do with how the engine sounds were transmitted to the CVR. However, they could not identify a more specific scenario during the initial phase of the investigation. ([18], p.176)

5.4 The First Round of Rudder PCU Testing

At the crash site, the tail of Flight 427 was the largest piece of wreckage. The rest of the plane had been squashed and shattered upon impact, but the tail was relatively intact.

Moreover, the tail contained the least-damaged hydraulic components, the **main rudder power control unit** (main rudder PCU) and the **standby rudder power control unit** (standby rudder PCU). ([18], p.114)

A rudder PCU is a hydraulic device that controls the rudder's movements. It takes commands from the cockpit through the rudder cables, and from the yaw damper[16] through electrical signals. It then directs hydraulic flows to push the rudder left or right. There are two rudder PCUs in the tail of a Boeing 737: The main rudder PCU powered by two main hydraulic systems A and B, and a standby rudder PCU that provides backup in case both systems A and B have failed. The input mechanisms to the two rudder PCUs are linked together at a mechanical joint called the torque tube. Even though the standby PCU is not pressurized in normal operations, it still moves in tandem with the main PCU. ([18], p.118)

Inside the rudder PCU, there is a soda can-sized component called the **dual concentric servo valve**. The phrase "dual-concentric" refers to the fact that the servo valve contains two tightly fitting cylinders (called the "slides"), one inside the other. The outer slide (called the **secondary slide**) is hollow, allowing the inner slide (called the **primary slide**) to move within it. The outer slide has holes bored through it that can match up with holes in the inner slide, and the servo valve housing that contains both slides also has holes in its wall. When the rudder is in neutral (not being operated), the slides block the holes in the wall of the servo valve housing, and the hydraulic fluid cannot pass through. When the slides move due to a pilot command or a yaw damper input, the holes line up to allow a flow of pressurized hydraulic fluid to go through. The fluid enters an actuating cylinder and moves the piston connected to the rudder. As a result, the rudder moves. ([18], p.120)

During the United Flight 585 investigation, the NTSB discovered a few possible failure mechanisms of the rudder PCU. First, when the investigators examined the standby rudder PCU recovered from flight 585, they found evidence that a rotating component of

---

[16] A yaw damper is a device on some aircraft that automatically reduces the rolling and yawing oscillations in turbulence.

its mechanical input arm, called the input shaft, had jammed against the bearing through which the shaft is supposed to rotate. The jamming seemed severe enough that galling, the transfer of materials between adjacent surfaces, had occurred between the input shaft and its bearing. Since the input arm to the standby PCU mechanically links to the input arm to the main PCU, the jamming to the standby PCU input arm could potentially influence the operation of the main rudder PCU. ([9], p.50)

Second, after the so-called Mack Moore incident in July of 1992, the NTSB discovered that the main rudder PCU recovered from the incident airplane exhibited an abnormal behavior called the **rudder reversal**. A rudder reversal occurs when the rudder moves in the opposite direction of the pilot command. Such behavior was possible in the Mack Moore airplane because of a manufacturing defect in the main rudder PCU dual concentric servo valve. Because the input lever that moved the secondary slide (the outer slide within the servo valve) could not maintain contact with its external stop, it moved the secondary slide too far beyond its designed operating position. The secondary slide's over-travel allowed the wrong holes on the servo valve housing to open, causing an abnormal hydraulic flow in the opposite direction of the intended one. ([9], p.69)

The investigators of the flight 585 crash eventually concluded that neither failure mechanism could account for the crash. However, the NTSB was sufficiently troubled by these failure mechanisms that it recommended the FAA to require the airlines to check for them regularly. ([12], p.104) After the USAir 427 accident, one of the first questions the investigators wanted to address was whether similar rudder failure mechanisms had played a role in the 427 crash.

Under the supervision of personnel from Parker Hannifin (manufacturer of the main rudder PCU from the accident airplane), the Systems Group investigators removed the two rudder PCUs from the wreckage and documented their conditions. On September 19 (11 days after the accident), the Systems Group arrived at Boeing's Equipment Quality Analysis Laboratory at Renton, Washington. There, guided by the Boeing engineers, they conducted metallurgical examinations of the standby rudder PCU input shaft. The investigators found signs of galling between the input shaft and its bearing, so they requested further tests on the effects of input shaft jamming on the rudder. ([10], p.68)

Two days later, the Systems Group arrived at Parker Hannifin's plant in Irvine, California, to test the main rudder PCU for any jamming or reversal. Although its input linkages were damaged, and the rod leading from the actuator piston (the unit that directly moves the rudder panel) was bent, the main body of the PCU was in good condition. The investigators found some hydraulic fluid remained in the dual concentric servo valve, with small metallic particles floating in it. Examination of the particles showed that they were flakes of nickel aluminum bronze alloy, a material commonly used for bearings. The investigators wanted to know whether the fluid's contamination level was high enough to jam the slides within the servo valve, so they took samples of the hydraulic fluid and sent them to Monsanto (the fluid's manufacturer) for testing. ([36])

The investigators further reasoned that if a metallic chip or particle had jammed the servo valve, it might have left witness marks on the slides or the interior of the servo valve housing. So after emptying the main rudder PCU of the remaining hydraulic fluid, they removed the slides from the PCU servo valve and inserted a borescope inside the device. The scope had a microscope with a prism at one end, allowing the investigators to examine the servo valve's interior. They found that the interior of the servo valve housing was free of witness marks. Similarly, microscopic examination of the servo valve slides showed only normal wear. ([18], p.122) To determine whether a chip or particle could have jammed the servo valve without leaving any witness marks, the Systems Group requested that Boeing conduct further tests.

In addition to examining its physical condition, the Systems Group also performed tests at Parker's facility to gauge the freedom of movements of the PCU's parts. The investigators replaced the mangled input linkages and the bent actuator piston rod with new ones. They then pressurized the PCU with new hydraulic fluid to see if it would jam or reverse, but it responded normally. When the investigators simulated the effect of yaw damper commands, the PCU also passed all the important yaw damper tests. ([18], p.122)

The tests requested by the Systems Group also failed to connect the rudder PCUs with the accident. First, tests on a replacement standby rudder PCU with galling patterns similar to those found on the accident airplane showed that the jamming of the standby input shaft could be (relatively) easily countered by pilot actions. When the test pilots applied 80 to 100 pounds of leg push against the rudder pedal (which they had no difficulty doing), they could override the effect of the jamming of the standby rudder PCU input shaft. ([10], p.70)

Second, tests of the sample hydraulic fluid extracted from the main rudder PCU showed a higher number and larger particles than the contamination level recommended by the manufacturer. However, the sample came from damaged components after the accident. The investigators could not be sure that fluid samples represented the true contamination level of the hydraulic fluid in the system before the accident. ([10], p.74)

Third, Boeing conducted a series of chip shearing tests at the NTSB's request in January, 1995. The investigators created chips of various materials found in the aircraft system. They inserted the chips into the holes at the interface between the primary (inner) slide and the secondary (outer) slide, and then moved the primary slide to close the holes. The tests showed that nearly all the chips sheared when a force up to 44 pounds was applied to the input lever of the primary slide, and only one type of chips made from hardened steel jammed the slides. Moreover, when the investigators examined the jammed servo valve, they found a physical mark on the primary slide where the chip was inserted, with size and shape similar to the inserted steel chip. ([10], p.76) In contrast, no such witness marks were found inside the servo valve from the main rudder PCU of the accident airplane. ([10], p.73)

In sum, the initial rudder PCU tests failed to identify mechanisms that could have jammed or reversed the main rudder PCU from the accident airplane without leaving physical traces. Moreover, the tests showed that even if the standby rudder PCU input shaft had jammed, it would not have caused the accident aircraft to lose control.

5.5 Ruling Out the Alternatives

With no promising leads emerging from the initial round of rudder PCU testing, The NTSB moved on to examine alternative scenarios, using two main tools. First, the investigators used a flight simulator to test different engine or flight control failure scenarios to see if any of them resulted in a flight path that matched the flight seconds of Flight 427. Second, the NTSB ordered a reconstruction of Flight 427's wreckage to see if there were traces characteristic of specific scenarios such as bombing, fuel tank explosion, and bird strike.

5.5.1 The M-Cab Flight Simulations

On September 22, 1994, at Boeing's multi-purpose cab (M-Cab) simulator facility in Seattle, the NTSB Performance Group attempted to replicate Flight 427's final flight path by putting the simulator through 45 different failure scenarios. ([18], p.129) Since Flight 427 experienced a sustained left roll and heading change, the investigators focused on failure scenarios involving asymmetric failures of flight controls or system components, such as the failure of one engine, the thrust reverser[17] being deployed only on one engine in flight, and uncommanded asymmetric activation of ailerons or spoilers[18] The candidate scenarios also included **rudder hardover**, the sustained deflection of the rudder (to the left in this case) at its full travel position (called the **blowdown limit**[19]). ([10], p.59)

The results of the flight simulations were inconclusive. First, most of the simulated scenarios failed to match the accident airplane's FDR data. For instance, at the engine power settings recorded by the FDR during the accident, an asymmetric thrust reverser

---

[17] A thrust reverser temporarily diverts the direction of the exhaust stream of an engine forward, so that the aircraft decelerates.

[18] Ailerons are panels near wingtips that move up and down, causing the lift on the wing to increase or decrease. The pilots move the ailerons by turning the control wheel. Spoilers are small panels on the top portions of the wing that decrease lift. Both control surfaces allow the pilots to roll the airplane or return to wings level.

[19] The blowdown limit represents a balance between the aerodynamic forces acting on the rudder and the mechanical forces produced by the PCU. On the ground, the blowdown limit of a Boeing 737-300 is ±26 degrees. During the flight, the blowdown limit is reduced depending on the airspeed and aircraft configurations.

deployment would not have produced a heading change that would match the FDR heading data. It would have produced longitudinal acceleration signatures not found in the FDR longitudinal acceleration data. Similarly, even the most severe asymmetrical aileron or spoiler extension could not produce a heading change rate of the magnitude recorded by the accident airplane's FDR. ([10], p.242)

Second, of all the simulations conducted, only the rudder hardover simulation produced results consistent with the FDR data from Flight 427. In particular, some of the simulations of the rudder hardover scenario produced a heading change rate similar to the FDR heading data recorded a few seconds before impact. However, because the test pilots who flew the M-Cab simulator responded differently to the rudder hardover scenario, the simulator results were not consistent among the pilots. Thus it was impossible to match the simulator results with the FDR data precisely. ([10], p.59)

Despite the inconclusive simulator results, the similarity between some results of the rudder hardover simulation and the FDR heading data prompted the NTSB to continue its investigation of rudder hardover scenarios. Moreover, to remove the individual variations introduced by pilots who participated in the M-Cab study, the NTSB and Boeing began to develop flight simulation programs on computer workstations. The workstation simulations would allow the engineers to manipulate the parameters of the rudder hardover scenario and the pilot responses to determine the details of the scenario that best matched the FDR data. ([10], p.59)

5.5.2 Wreckage Reconstruction

Following the M-Cab simulator study, the NTSB reconstructed the wings and the fuselage of the accident airplane between October 30 and November 11, 1994. Part of the reconstruction's purpose was to evaluate several failure scenarios that were difficult to simulate in the M-Cab simulator, including inflight fire, fuel tank explosion, bomb, collision with birds or other airborne objects. Each of these possible scenarios would leave behind distinctive traces, and a reconstruction of the wreckage could help the investigators find these traces. ([10], p.39)

Due to the ground impact, many pieces of the wreckage were too fragmented or severely damaged by post-impact fire to be identified. As a result, a three-dimensional reconstruction of the fuselage of Flight 427 was impossible. Instead, Flight 427's reconstruction was two dimensional. The investigators spread out full-scale drawings of the plane components on the hangar floor and laid pieces of the wreckage over their locations on the drawings. ([18], p.104) Because of its experience in reconstructing the Boeing 747 involved in the bombing of the TWA Flight 103 over Lockerbie, Scotland, the Air Accident Investigation Branch (AAIB) in the UK also participated in the reconstruction of Flight 427. ([10], p.39) Overall, the reconstruction focused on four main areas ([18], p.103):

First, the investigators reconstructed the radome (nose) of the aircraft, the forward pressure bulkhead[20] and the flight controls on the leading edge of the wings. This reconstruction helps evaluate whether there were any indications of a bird strike. The possibility of a bird strike was salient because witnesses reported large flocks of migratory birds in the Pittsburgh area throughout the afternoon and evening of the accident. ([10], p.42)

The investigators subjected the reconstructed components to ultraviolet light inspection for bloodstain and other traces of bird. Even though no evidence of bird remains was found on most of the examined areas, a small stain on the outer surface of a leading-edge slat[21] did exhibit an intense white fluorescence when illuminated by ultraviolet light. The investigators took two small samples of the fluorescent debris from the slat surface and sent them to an ornithologist at the Smithsonian Institution's bird division for examination. ([10], p.42) The examination's result was negative: The debris exhibited no characteristics that resembled those of a bird. ([10], p.43)

---

[20] The forward pressure bulkhead is the front wall of the cockpit below the windshield and in front of the pilot's legs. It is a barrier for the pressurized interior cabin.

[21] Slats are flight control panels on each wing's leading edge. When deployed, they allow the aircraft to fly at lower airspeed during takeoff and landing.

Second, the investigators reconstructed portions of the floor beam and the doors to see if they could have failed before the crash. This reconstruction was conducted because of a few prior accidents: On June 12, 1972, American Airlines Flight 96 (a DC-10) flew above Windsor, Ontario, when its left rear cargo door blew open and fell off. The rapid decompression of the cargo bay caused a floor beam collapse, which jammed the rudder control cable and caused the rudder to deflect to its maximum right position (hardover). In this accident, the pilots managed to return to Detroit Metropolitan Airport and land safely without major injuries. However, on March 3, 1974, a nearly identical cargo door failure happened on the Turkish Airlines Flight 981, causing it to crash near Paris and killing all 346 people on board. In the light of these prior accidents, the NTSB wanted to know whether similar door failures or floor beam collapse had played a role in the Flight 427 crash. ([18], p.103)

The floor beam structures of Flight 427 were severely fragmented, and the amount of identifiable floor beam structure varied at each fuselage location. Forward the wings, less than 5 percent of the beams could be identified. Things improved towards the tail, where a maximum of 95 percent of floor beams could be identified. Overall, only about 50 percent of the beams were identified in the reconstruction. ([10], p.39) The identified floor beams showed no signs of pre-impact, fire, heat, or explosives damage. Other than that, very little could be inferred about their conditions before the crash. The investigators had clearer results concerning the doors: they identified enough locking mechanisms and door frame pieces to conclude that all the doors were in closed positions at impact. ([10], p.40)

Third, another main area of reconstruction was the auxiliary long-range fuel tank. This fuel tank was in the cargo bay and had two possible failure modes relevant to the 427 accident. First, the fuel vapors in the nearly empty fuel tank might have exploded. Second, the fuel pump might have been accidentally switched on and left running, creating a vacuum in the tank and eventually causing the tank to collapse. ([18], p.103) Either way, there was a possibility that the floor beams above the fuel tank were affected, leading to the severing of the hydraulic lines or control cables to the rudder.

The investigators were able to recover and identify about 40 percent of the fuel tank structure and over 50 percent of its electrical control components. The identified fuel tank components did not show any pre-impact fire or heat damage or any other abnormal characteristics. The positions of the fuel tanks valve were consistent with the fuel tank being deactivated. ([10], p.41) Moreover, Boeing's drawings showed that the tank was not bolted to the floor beams above it, meaning that the possibility of the floor rupturing due to a fuel tank collapse was remote. ([18], p.104)

Finally, the last main area of wreckage reconstruction was the main landing gear wheel-wells. Even though it was unlikely in the Flight 427 accident, an overheated brake could have caused a fire in the wheel well. Similarly, if a rapidly spinning wheel disintegrated when it was retracted into the wheel well, it could cause structural damage and perhaps rupture hydraulic lines or control cables. To evaluate these possibilities, the investigators looked for a splattering of melted rudder in the wheel well, which might indicate a tire overheated or on fire. However, the examination of the tires, wheels, and the wheel wells showed no signs of pre-impact failure. ([18], p.104)

Based on the wreckage reconstruction, the NTSB concluded that USAir flight 427 did not experience a fuel tank explosion, bomb, bird strike, or structural failure. Also, the NTSB was able to rule out various other possible scenarios such as severe atmospheric phenomena or collision with another aircraft: Unlike the Flight 585 accident, the weather in the Pittsburgh area was clear with light winds at the time of the accident. Moreover, the ATC radar data showed no nearby aircraft that could have collided with Flight 427. ([10], p.241)

5.6 Wake Turbulence

As more and more alternative scenarios were ruled out, the few remaining possibilities were examined more closely. Other than the rudder hardover, the other main remaining possibility was **wake turbulence**, the turbulence *generated* by another aircraft in flight.

Very little was known about wake turbulence until the FAA and NASA performed a series of tests following three accidents caused by wake turbulence between 1964 and 1972. The tests showed that the worst part of wake turbulence came from the aircraft's wingtips. When air flew over each wingtip, it spun off in a vortex called wake vortex. Moreover, while the wake vortices tend to sank vertically through the air and decay over time, they could persist at an altitude in calm air for 2-3 minutes. When a smaller, lighter aircraft flew into the wake vortices generated by a larger, heavier aircraft, it could get rolled about or even flipped over. ([18], p.128) Because of these discoveries, the FAA increased the minimum aircraft separation standards to ensure that an aircraft, especially a smaller one, is not endangered by the effects of wake vortices generated by a preceding aircraft. ([10], p.244)

Was there a larger aircraft in Flight 427's vicinity at the time of the crash? To answer this question, the Air Traffic Control (ATC) Group searched the departure and arrival records at Pittsburgh International Airport (PIT) on the day of the accident. The ATC records showed that there were only two aircraft operating in the neighborhood of flight 427: (1) Atlantic Coast flight 6425, a Jetstream commuter flight that had just departed PIT and was climbing, and (2) Delta flight 1083, a Boeing 727-200 that preceded USAir flight 427 to PIT. ([10], p.54) A Boeing 727 is only slightly larger than a Boeing 737, and the Jetstream aircraft was smaller. There were no significantly larger aircraft such as a Boeing 747 or a DC-10 nearby. ([18], p.96)

To determine whether the wake vortices from either the Atlantic Coast or Delta airplanes might have played a role in the USAir 427's accident sequence, the Performance Group investigators plotted the radar tracking data for these three airplanes on the U.S Geological

Survey Map. ([10], p.54) The radar data indicated that Atlantic Coast flight 6425 departed PIT and headed north, whereas USAir 427 approached PIT and headed east. The two aircraft's radar tracks did not cross at any time, meaning that the Atlantic Coast aircraft was unlikely to be a factor in the accident. ([10], p.54)

The radar track of Delta flight 1083 was more interesting. Delta flight 1083 was also en route to Pittsburgh, and it had followed the same track that Flight 427 was maneuvering to join when the accident sequence began. The radar data indicated that Delta flight 1083 descended through 6,300 feet and headed east when it passed the horizontal location where the initial upset of USAir flight 427 subsequently occurred. When flight 427 reached that horizontal location about 69 seconds later at an altitude of about 6,000 feet, the failure sequence began. At that time, the distance between the two airplanes was 4.5 nautical miles, larger than the minimum separation required by the FAA. ([10], p.55) Nevertheless, the fact that the two radar tracks crossed almost the exact horizontal location where the accident sequence initiated was too much of a coincidence to ignore. It raised a further question: Did flight 427 encountered the wake vortices generated by Delta flight 1083 at the time of its initial upset? ([18], p.128)

The aerodynamics experts from the NTSB and NASA studied the most likely movement of the wake vortices produced by Delta flight 1083, given the airplane's estimated weight and flight control configuration. The study indicated that the wake vortices produced by the Delta flight would have descended at 300-500 feet per minute. Given these rates, the wake vortices would likely have descended from 6,300 feet to about 5,800-6,000 feet during the 69 seconds after the Delta flight passed the initial upset location. Since USAir flight 427 was at an altitude of 6,000 feet near the initial upset location, it follows that USAir flight 427 likely encountered the wake vortices produced by Delta flight 1083 at the time of the initial upset. ([10], p.55)

The next question then became: Could the wake vortices from Delta flight 1083 have caused the drastic yawing and rolling movement experienced by USAir flight 427? Delta flight 1083 was a Boeing 727, which has a similar wingspan as the USAir Boeing 737 and was not regarded as dangerous as far as its wake turbulence was concerned. ([18], p.128) Nevertheless, the investigators did not have data on the effects of a Boeing 737 entering the wake vortices of a Boeing 727. They were unsure about how severe the effects might be.

In mid-October, 1994, the NTSB Performance Group returned to Boeing's M-Cab facility at Renton. This time, they programmed the simulator to replicate the effects of Flight 427 entering the wake vortices of an airplane about 4 miles ahead. Moreover, the investigators used the same computer program developed three years ago to simulate a mountain rotor's impact on Flight 585 at Colorado Springs. Rotors are much larger than wake vortices, but the program was still useful because it could vary the strength and diameter of the vortex's core and change the direction of its rotation. The investigators added two counter-rotating rotors to the program to represent wake vortices, varied the core sizes between 4 feet and 17 feet, and separated the two vortices by 85 feet to represent the wingspan of a Boeing 727. When test pilots from the Performance Group flew in the M-Cab simulator, they agreed that the computer program gave a realistic representation of wake vortices. ([18], p.130)

The test pilots flew the simulator into the wake vortices at varying intercept angles to see the effects. They found that even at higher core strength, wake vortices had relatively minor effects on aircraft controllability. Depending on the angle of the entrance, the simulator could roll between 10 degrees and 30 degrees. However, the simulated aircraft would stay in the vortex for about 5 seconds and stabilize soon after ejection from the vortex. Invariably, the pilots were able to regain control easily after the initial roll. ([18], p.139) Assuming that the wake vortex simulations were accurate, the wake vortices might have initiated the failure sequence. However, they could not have been responsible for its continuation because the aircraft would have been in the vortex in just 5 seconds. Unless there were serious inaccuracies in the simulation program, something else must have sustained the continued left roll and yaw of Flight 427. ([18], p.130)

5.7 Human Performance

The lack of plausible alternative scenarios further supported what the NTSB investigators had suspected all along: That only a rudder hardover—the sustained deflection of a rudder at its full travel position—could have caused the drastic rolling and yawing of USAir flight 427. However, the rudder hardover scenario gave rise to a further question: If a rudder hardover had occurred in the accident airplane, what had caused it?

A rudder hardover is an extreme event and could cause an aircraft to yaw and roll violently. Because of the powerful effects of the rudder, the 737 pilots may use quick rudder movements to control the aircraft in difficult situations. However, they seldom use the rudder to steer the plane in normal flight. Instead, the pilots mainly used the ailerons on the wings to turn the airplane. An automatic device called the yaw damper makes minor adjustments of the rudder (a maximum of ±3 degrees of rudder deflection) to smooth the flight in turbulence. ([18], p.12) Nothing in a normal flight would cause the rudder to deflect at its blowdown limit for an extended time. If a rudder hardover had occurred in USAir 427, it must have been caused either by a mechanical rudder system anomaly or by abnormal pilot actions. ([10], p.245)

Until February 1995, the NTSB had focused on examining the possibility of a rudder hardover caused by some rudder system anomaly such as the jamming or reversal of rudder PCU. After the initial rudder PCU tests failed to detect any anomaly that could have contributed to the accident, one of the major parties in the NTSB investigation—Boeing—began to push the NTSB to examine the possibility of a rudder hardover caused by abnormal pilot actions.

The NTSB and Boeing had a delicate relationship in the investigation of the USAir 427 accident. The NTSB subscribed to the so-called party system. Any organization implicated in the accident could apply to be affiliated to the investigation as a party. The parties participate in the probe, and can even suggest avenues of further investigations. A major upside of the party system is that the NTSB could draw on the participating parties' expertise and resources. ([18], p.xvii) Due to their technical expertise, Boeing employees constituted a large percentage of the investigators in the USAir 427 case. For instance, of the 24 people who would eventually make up the Structures Group, 11 were from Boeing. ([10], p.81) On the other hand, Boeing's self-interests were at stake in the USAir 427 investigation, and it was motivated to exert influence on the investigation when it could.

In mid-February, 1995, Boeing's director of flight safety wrote to the NTSB's investigator-in-charge that a full Human Performance Group should be formed for the USAir 427 accident. The NTSB had had a human performance subgroup within its Operations Group earlier in the investigation. The investigators had found nothing out of the ordinary in the background of the pilots. However, Boeing wanted the NTSB to probe a more specific possibility: That the pilots had incorrectly pressed the left rudder pedal throughout the accident sequence, causing a rudder hardover to the left. ([18], p.147)

After the NTSB followed Boeing's suggestion and formed a new Human Performance Group, the Boeing representatives continued to argue for the need to pursue the possibility of incorrect pilot inputs to the left rudder pedal. Boeing cited psychological studies of cognitive errors and cases in which the pilots pressed the wrong rudder pedal. ([18], p.156)

First, Boeing suggested the 427 flight crew might have responded to the unexpected encounter with significant wake turbulence by incorrectly applying the left rudder. To motivate this suggestion, Boeing psychologists cited two main studies. One study was done by the Royal Air Force's Institute of Aviation Medicine. It examined 148 crashes from 1972 to 1988 and found that a significant portion of the crashes involved cognitive failures due to under-arousal or over-arousal. The other study was a lab experiment by FAA's Civil Aerospace Medical Institute. According to this study, subjects who had been unexpectedly startled by a sudden loud noise took longer to complete simple tasks. Boeing used these two studies as evidence that the pilots could have reacted improperly to being startled by making the incorrect left pedal input. ([18], p.158)

Second, Boeing psychologists found two cases in which the pilots' medical conditions led to incorrect application of the rudder pedal, one in 1980 and the other in 1994. The two incidents share the same storyline: The aircraft was on the final approach to its destination airport when the first officer (the pilot flying) suffered from a seizure. As a result, the first officer's left leg extended and unintentionally pressed down on the rudder pedal. The captain initially struggled with countering the roll of the aircraft caused by the rudder. After a flight attendant moved the first officer's leg off the rudder pedal, the captain regained control of the airplane. ([10], p.183) In both incidents, the captain admitted that they had been startled, which had delayed their comprehension of the situation. ([18], p.157)

Third, another suggestion of Boeing was that when the aircraft started its roll due to wake turbulence, the pilots became confused through a process known as spatial disorientation. Boeing cited a book on flight deck performance, according to which spatial disorientation could contribute to incorrect pilot control inputs. The book stated that pilots use vestibular (inner ear) and visual (both the horizon and flight instruments) cues to determine the airplane's position in space. When an aircraft turns or accelerates rapidly, the inner ear may give a false signal, which, combined with the lack of visual cues, may cause the pilots to be disoriented about their spatial orientation. ([10], p.184)

Based on its suggested scenarios, Boeing raised the question about the USA 427 crew's medical records and proposed that the Human Performance Group check them to see if either pilot had any medical conditions relevant to the accident. In addition, Boeing proposed to examine the pilot's training records and see if their instructors had noticed any predisposition to misapply the flight controls. Other proposals by Boeing included further validating the wake turbulence encounter scenario and re-examining the CVR tape to better understand the pilots' actions and emotions during the accident sequence. ([18], p.158)

Some participating parties of the investigation strongly objected to Boeing's proposed line of inquiry. USAir and ALAP (Air Line Pilots Association), particularly, had a stake in defending the USAir flight 427 pilots' reputation. Representatives from these parties argued that there was no evidence that the pilots had applied improper rudder input, and that the investigation should focus on identifying malfunctions of the rudder control system. ([18], p.149) Despite these protests, the NTSB took Boeing's suggestions seriously. The Human Performance Group reviewed the medical and training records of the captain and the first officer of USAir flight 427, but found nothing noteworthy. The training records and interviews with people acquainted with the flight crew indicated that both pilots were meticulous, highly skilled, and were able to stay calm during emergencies. The medical records showed that, other than a back surgery that the captain had undergone, both pilots had excellent health during the five years before the accident. ([10], p.9-10)

To test the pilot disorientation scenario, the Human Performance Group conducted simulations at NASA's Ames Research Center in California. NASA-Ames had the world's largest vertical motion simulator (VMS), which could reproduce high-G forces by rising and falling rapidly through as much as 60 feet. The investigators programmed the VMS to simulate Flight 427's encounter with wake turbulence, and they recreated a clear visual horizon on the simulator screen based on weather reports of the accident. During the simulations, a NASA scientist specializing in the effects of acceleration on spatial disorientation sat in one of the simulator seats, and a Human Performance Group test pilot occupied the other seat. ([10], p.185)

After the simulator tests, the NASA scientist described his impressions of the wake turbulence encounter as follows:

> I was surprised at how gentle it all was. I had thought that the upset would be more severe. It was a surprise, it did get my attention. But it was not a violent kind of an upset that would have me fail to know where I was and what my orientation was. ([10], p.185)

The NASA scientist strongly believed that the USAir 427 pilots did not experience spatial disorientation. There were clear external visual cues (the sky, ground, and horizon), and the motions of the wake turbulence encounter were not excessively violent. The scientist further stated that, during post-simulator-ride interviews, the simulation pilots reported that they always knew the airplane's location and orientation and that they could have flown out of the wake turbulence portion of the upset event. Based on these judgments, the NTSB concluded that pilot disorientation was not a causal factor in the accident. ([10], p.185)

5.8 The Simulator Validation Tests

In September and October 1995, the NTSB conducted two sets of flight tests, one in Seattle and one in Atlantic City, New Jersey. The flight tests were called the simulator validation tests because they were supposed to validate existing and acquire additional aerodynamic data for Boeing's M-Cab simulators. Much of the investigation was based on M-Cab's computer estimates about how 737s would behave. Some participating parties in the investigation (notably the FAA) wanted to be sure that M-Cab accurately simulated real 737s. ([1], p.144) The NTSB leased a Boeing 737-300 from USAir and used it in both sets of flight tests. ([10], p.63)

5.8.1 The Seattle Flight Tests

The Seattle flight tests began on September 20, 1995. The purpose of the tests was to measure the 737's responses to various flight control inputs that induced roll and yaw. Also, the tests gave the investigators a chance to learn more about the **crossover airspeed**, the airspeed at which the yaw and roll effects of a rudder hardover exceed the maximum roll control provided by the control wheel. ([10], p.63)

When the first generation Boeing 737-100 was certified in 1967, Boeing informed the FAA that if a failure in the rudder control system caused a sudden, uncommanded movement of the rudder, the pilots could counter the yaw and roll by turning the control wheel in the other way and deflecting the ailerons on the wings. Boeing later discovered that this was not true at lower speeds. However, the company did not regard the discovery as critical and did not mention crossover airspeed in its flight manuals or alert airlines. Members of the NTSB Performance Group had noticed the crossover airspeed during the M-Cab tests in 1994, but they needed data from a real 737 to determine the speed. That information could also help to answer the question about whether the pilots could have prevented the crash. ([1], p.145)

During the Seattle flight tests, several test conditions examined whether the test pilots could maintain control of the airplane and a constant heading using the control wheel to oppose full rudder deflections. The test pilots set the flaps to flaps one, the same configuration in Flight 427. They then pushed the left rudder pedal and simultaneously turned the control wheel to the right to keep the plane from rolling, a maneuver known as a steady-heading sideslip. ([1], p.145)

The tests showed that given the flaps one configuration and USAir 427's weight, the crossover airspeed was 187 knots, significantly higher than the investigators expected. At airspeed above 187 knots, control wheel input could counter the roll induced by a full rudder deflection. However, at or below 187 knots, the airplane continued to roll in the direction of the rudder deflection, despite full control wheel input in the opposite direction. ([10], p.63) Successful recovery required the test pilots to immediately pitch the aircraft nose down to maintain a speed above the crossover point, and to apply full control wheel input in the opposite direction at the same time. ([10], p.64)

The discovery of USAir flight 427's crossover airspeed was significant because it was very close to flight 427's airspeed at the beginning of the accident sequence (about 190 knots). In other words, flight 427 was right at its crossover point when it began to lose control. ([1], p.145)

5.8.2 The Atlantic City Flight Tests

On September 25, 1995, the NTSB began a series of flight tests near Atlantic City to examine the aerodynamic effects of 727-generated wake vortices on a 737. The investigators borrowed a Boeing 727 from the FAA and equipped it with wing-tip smoke generators to help visualize the wake vortex core. They also equipped the Boeing 737 borrowed from USAir with an enhanced FDR, a quick-access cockpit voice recorder, and seven video cameras (including one on the tail). An observation airplane provided by Boeing also carried video recording equipment to document the flight tests and the weather conditions. ([10], p.55)

During the flight tests, the 737 penetrated the 727's wake vortex cores (indicated by the wing tip smoke) about 150 times from various intercept angles, at various altitudes, and separation distances of 2 to 4.2 nautical miles.[22] For other flight test conditions, the flight test pilots maneuvered the 737 so that specific airplane surfaces (including the wings, vertical fin, and the engines) passed through the wake vortex cores. After the flight tests, the investigators collected information from the tapes of the enhanced FDR, the video cameras, the quick-access CVR installed on the 737, and test pilot statements. ([10], p.56)

The investigators compared the flight test data with the results of M-Cab simulations performed at Boeing. The comparison indicated that the M-Cab simulations adequately predicted the lift, roll, and pitch movements induced by wake vortices. The only thing that the simulation model did not accurately predict was the wake-induced yawing movement characteristics. The flight tests' videotapes showed that, when the airplane passed over or directly through the wake vortex cores, the 737's wings and fuselage disrupted the wake vortex. Under these circumstances, the yawing motions experienced in the flight tests were significantly less than predicted by the simulations. However, when the airplane was slightly underneath the wake so that its vertical tail surface passed through the wake vortex core, the wake vortex that contacted the vertical tail surface had not been previously disrupted.

---

[22] USAir flight 427 and Delta flight 1083 were 4.5 nautical miles apart at the initial upset.

This was the only situation in which the yawing movements produced in the test flights exceeded the yaw predicted by the M-Cab simulation model. ([10], p.58)

The statements of the flight test pilots also supported the results of the M-Cab simulations. According to the flight test pilots, although the wake encounter could produce rolls between 10 and 30 degrees, the effects usually lasted only a few seconds. When the pilots used flight controls to counteract the wake-induced roll, they mostly used the ailerons and rarely used the rudder. The pilots agreed that they did not experience anything during the tests that would result in a loss of control, or prompt a pilot to apply and hold full rudder. ([10], p.57) Most of the flight test pilots described the wake encounters as "routine" and not startling. However, the test pilots from Boeing stated that a strong wake encounter would likely be startling to pilots when encountered during an otherwise smooth flight. ([10], p.58)

Based on the flight test data about the wake turbulence encounter, Boeing refined its wake vortex simulation model used in the M-Cab, and the Performance Group ran additional flight simulations of the effects of Delta flight 1083's wake vortices on USAir flight 427. The simulations further confirmed that encounters with wake vortices did not result in significant control problems. ([10], p.58)

5.9 CVR Sound Analysis

The biggest breakthrough of the simulator validation flight tests was the identification of a few mysterious sounds recorded on Flight 427's cockpit voice recorder tape. The NTSB acoustic experts had installed a quick-access cockpit voice recorder on the test airplane rented from USAir. By analyzing the sounds recorded during the flight tests, they were able to identify the causes of two sets of sounds recorded at the beginning of flight 427's accident sequence: A series of thumping sounds, and two instances of the engine increasing in loudness.[23]

---

[23] The investigators were not able to identify all the sounds recorded at the beginning of the accident sequence. For instance, the CVR recorded two "clickety-click" sounds after the initial upset, and the

5.9.1 The Thumps

During the examination of the CVR recording from USAir flight 427, the Cockpit Voice Recorder Group investigators noticed a few sounds that they characterized as “thumps” at the beginning of the accident sequence. At 19:02:56.5, the accident airplane’s CVR recorded three thumps within 1 second, followed by two more thumps between 19:02:58 and 19:02:59. Examination of the entire 31-minute CVR tape showed that the CVR had not recorded anything like the thumps during the previous 30 and 1/2 minutes. The investigators were initially unable to determine what caused these thumps. ([10], p.130)

The investigators noticed, however, that the thumps were recorded by two different channels of the CVR: The cockpit area microphone (CAM) in the cockpit ceiling, and the microphone located at the jump seat[24]. This suggested that the thumps might have been transmitted to these microphones through the air and the airplane’s metal fuselage structure. Since sound travels much faster in metals than in the air, it was possible to calculate the approximate distance of the source of the thumps based on the time gap between the sound transmitted through metal and the sound transmitted through the air. Moreover, the sound signals would arrive first at the microphone closest to the source. So it was possible to calculate the approximate direction of the source of the thumps based on the time gap between the sound signals recorded on the two microphones. ([10], p.131)

In September 1995 (before the wake turbulence flight tests), the CVR Group conducted a test to determine the location of the source of the thumps. The investigators configured a test 737 to represent the accident airplane’s condition at the time the thumps were recorded. They started the CVR in the test airplane, then struck the airplane structure at various locations (both inside and outside) with a rubber mallet, carefully noting each strike’s time and location. Back in the NTSB lab, the investigators measured the gap between the time of the airborne sound and the faster fuselage-transmitted sound for each source location.

---

investigators were unable to identify their sources. ([10], p.135)

[24] A jump seat is an auxiliary seat in the cockpit for flight personnel not operating the aircraft.

The fuselage-transmitted sound was easily distinguished from the airborne sound because the former has a characteristic lower frequency. ([18], p.175)

The investigators then scrutinized the thumps' timing on the original Flight 427 CVR tape, using waveform printouts to accurately place the onset of both the fuselage-transmitted and the airborne sounds. When they compare the time gap in the accident airplane and the test airplane, they discovered that the best match was a sound from a source approximately 12 to 16 feet back from the cockpit area microphone (CAM), corresponding to a location near the No.1 cabin door and row 1-2 in the first class. This settled the question about (roughly) where the the source of the sound was. ([18], p.175)

However, even though the rubber mallet strike tests could duplicate the thumps' timing, the sound signatures produced during the tests were distinctly different from those of the thumps. The question remained: What had caused the thumps?

After the Atlantic City wake turbulence flight tests, the test pilots reported that when the main fuselage passed through the center of the wake core, they heard "whooshing" sounds in the cockpit. When the CVR acoustic experts asked the test pilots whether the "whooshing" sounds resembled the thumps recorded in the accident CVR, the pilots said they did not match. ([18], p.173) However, when the investigators compared the sounds recorded during the wake turbulence tests with the CVR sounds from USAir flight 427, they found that the sounds recorded in wake encounters had signatures very similar to the thumps. ([10], p.131)

To verify that wake vortices caused the thumps, the CVR Group correlated the videotapes shot from the test airplane and the observation airplane with the tapes of the cockpit sound recording from the test airplane. Then, frame by frame, they moved the video and sound tapes forward until they reached the point when the thumps started. The videotapes showed that the fuselage was entering a wake vortex illuminated by smoke trails at that time. On one side of the test airplane (facing the inside turn), there was a steady stream of smoke along the fuselage. On the other side, however, the smoke did not flow smoothly.

Instead, the smoke separated from the fuselage for a few feet and reattached to the fuselage for the remainder of its length. The point where it rejoined the fuselage was the same location where the investigators had determined the thumps originated. ([18], p.176)

The CVR Group's findings confirmed that the wake vortices of the Delta airplane caused the thumps. They also explained why the test pilots said the "whooshing" sounds they heard in the cockpit differed from the thumps recorded on the accident CVR. The whooshing noise the pilots heard was an airborne sound, which had a higher frequency. The thumps, however, originated on the fuselage structure and were recorded by the cockpit area microphone (CAM) attached to the cockpit ceiling. As a result, the thumps were combinations of high-frequency sounds transmitted through the air and lower-frequency sounds transmitted by the airplane structure. Hence they would sound differently from airborne sounds alone. ([18], p.176)

5.9.2 The Engine Loudness Increase

Another set of mysterious sounds recorded by the CVR at the beginning of USAir flight 427's accident sequence was two instances of an increase in engine loudness. A sound spectrum study of the CVR recording showed that at 19:02:58.27 and 19:03:02.3, the sounds associated with both engines simultaneously increased in volume by about 30 percent. At the same time, the engine sounds' frequency, and the sound signatures of all other background noises remained constant. ([10], p.132) The increases in engine loudness baffled the NTSB investigators because they did not correspond to the engine power setting recorded by the FDR at the time. Moreover, the NTSB could not identify a mechanism that could increase the CVR's ability to record the volume of the engine sounds, while keeping the sound signatures from all other frequencies unchanged. ([10], p.132)

Surprisingly, the CVR tape from the Seattle flight tests resolved the puzzle. During the Seattle tests, the pilots performed a maneuver known as the steady-heading sideslip, which involved deflecting the rudder in one direction and then countering it with the opposite aileron. When investigators from the CVR Group listened to these maneuvers' tapes, they were surprised to observe similar increases in engine volume in some of the maneuvers. Moreover, they discovered a correlation between the rate of rudder deflection and the increase in engine volume recorded in the CVR: Larger and more rapid rudder displacements resulted in greater engine volume changes. In contrast, more gentle rudder movements barely changed recorded engine volume at all. The engine sound signature that best matched the CVR recording from the accident airplane corresponded to a test condition where the rudder deflected rapidly from 0 to 14 degrees. ([10], p.133)

In short, a large and rapid deflection of the rudder caused the increase in engine volume in the accident CVR. *How* did rapid rudder deflection make a difference to the recorded engine sound volume? The NTSB did not have a conclusive answer to this question. However, an acoustic expert from the CVR group hypothesized the following mechanism: The engines of the Boeing 737-300 were designed to suppress side noises and improve passenger comfort, so they emit most of the noise out of the engine's front during normal flight. When the airplane yawed sharply, the sounds coming from the engines impinged differently onto the fuselage, causing a temporary increase in the engine sound volume recorded in the CVR. ([18], p.178)

5.10 The Independent Technical Advisory Panel

The simulator validation flight tests and the sound analysis of the test flight CVRs strengthened the NTSB's convictions in several key conclusions. First, USAir flight 427 had encountered the wake vortices of Delta flight 1083. Second, the wake vortices had contributed to the initial roll and yaw of the aircraft but were insufficient to cause the roll and yaw to continue after the first few seconds. Third, the rudder of the accident airplane had deflected rapidly to the left during the first few seconds of the accident sequence. Finally, a sustained rudder deflection at its blowdown limit (i.e., a rudder hardover) caused the accident airplane to continue to roll and yaw throughout the accident sequence.

However, the main question raised by this picture of the accident remained unresolved: What had caused the rudder hardover? Examinations of the two chief candidate answers to this question—a rudder system anomaly versus abnormal pilot actions—had failed to reveal any plausible mechanism that could have caused a rudder hardover in the accident airplane.

The NTSB had been more inclined towards the possibility of a rudder system malfunction, and its Systems Group investigators had conducted a variety of tests of two rudder PCUs from USAir flight 427. By the end of 1995, however, they had found nothing noteworthy: The clearances between the slides and the servo valve housing of the main rudder PCU were tighter than usual, but they still met Boeing's specifications. Dynamical testing of the main rudder PCU using new hydraulic fluid failed to produce a jam or reversal. The chip shearing tests showed that the main rudder PCU could still operate even if its hydraulic fluid were highly contaminated with debris. Moreover, metal chips could not jam the PCU without leaving characteristic witness marks, which were not found in the main rudder PCU from the accident airplane. As for the standby rudder PCU, the jamming of its input shaft could not produce enough rudder deflection to cause a hardover. ([10], p.69-76)

In January 1996, the Chairman of the NTSB announced that he would form an independent advisory panel to review the work by the Systems Group in the USAir flight 427 investigation. The advisory panel's role was to ensure that the Systems Group did not miss out on any relevant failure modes of the rudder control system and suggest further areas of study that the Systems Group should pursue. ([18], p.186) The NTSB and the FAA selected the six members of the advisory panel. They consisted of aircraft hydraulic systems experts from NASA, the FAA, the U.S Air Force, and two hydraulic component manufacturing companies. ([10], p.162)

On February 8, 1996, the independent technical advisory panel held its first meeting in Washington, D.C. During the meeting, a panel member described an experience he had at Bendix Corporation back in 1966. The company tested hydraulic servo valves similar in design to the rudder PCU servo valve from USAir flight 427, in preparation for a bid on Boeing's contract for servo valves in the then-new Boeing 747. The test servo valves failed a thermal shock test designed to check the valve's ability to tolerate the extreme difference between the temperature at cruising altitude and the hydraulic fluid overheated by hydraulic pumps. ([18], p.188) The thermal shock test showed that after the test servo valve was frozen to minus 40 degrees and then injected with heated hydraulic fluid, it would jam for a few seconds. As a result, the Bendix engineers had to redesign the servo valve to fix the problem. Since the rudder PCU servo valve in the Boeing 737 was designed during the same era by another company, the panel member theorized that they might have a similar problem. ([34])

Another panel member from the U.S Air Force also reported that he had worked on a military fighter project that had used a PCU servo valve similar to the 737 main rudder PCU servo valve. According to this panel member, there was a crash caused by a jammed PCU servo valve very early in the initial production test flights. The investigation of that accident showed that the servo valve jammed because hot hydraulic fluid had entered the cold valve body, causing the thermal expansion of the servo valve's inner slides into the servo valve housing. ([10] p.162)

Even though the panel also discussed other aspects of the Systems Group's work and other possibilities of a rudder system malfunction, the topic of thermal shock dominated the meeting. The panel members raised the question about whether Flight 427's main rudder PCU servo valve could have jammed due to a thermal shock, and made a proposal to conduct special tests to address this question. However, due to the NTSB's party system, the tests could not proceed without the approval of all the members of the Systems Group (including representatives from all the participating parties). The decision-making process was time-consuming, and months elapsed before all the parties agreed to the thermal shock tests' details. ([18], p.189-190)

5.11 The Eastwind 517 Incident

Before the thermal shock tests were carried out, however, an incident happened in the summer of 1996 and heightened the NTSB's focus on the 737 rudder control system. On June 9, 1996, at about 10 pm, Eastwind Airline flight 517, a Boeing 737-200, experienced uncommanded yaw and roll to the right near Richmond, Virginia. The airplane descended to an altitude of about 4,000 feet and was flying at an airspeed of 250 knots when the yaw and roll occurred, and one flight attendant suffered minor injuries during the upset. Fortunately, the pilots regained control of the airplane and landed at the destination airport without further injury to the aircraft occupants. ([10], p.51)

During interviews after the accident, the captain stated that he was manually flying the airplane with the autopilot switched off, which was his habit on approach. Also, his feet rested lightly on the rudder pedals. As the airplane descended through 5,000 feet, he felt a brief "kick" or "bump" on the right rudder pedal, although the pedal did not move. The captain glanced at the first officer to see if he had pressed the pedals, but the first officer had his feet on the floor. ([10], p.51)

As the airplane descended through 4,000 feet, however, the airplane yawed abruptly to the right and rolled to the right. The FDR on the incident airplane indicated that the airplane rolled to the right about 10 degrees, with a simultaneous heading change of about 5 degrees per second. ([10], p.52) The captain stated that he immediately countered the movement by pressing hard on the left rudder pedal and turning the control wheel to apply the left aileron. ([10], p.51) The captain further stated that the rudder pedal felt stiffer than normal and did not seem to respond to his input. The first officer described the captain as "fighting, trying to regain control" and "standing on the left rudder (pedal)". ([10], p.52)

According to the captain, the airplane was still trying to roll despite his rudder and aileron inputs, so he advanced the right engine throttle to counter the right roll further. The three countermeasures helped to swing the airplane back toward level flight and even caused the airplane to bank left momentarily. However, the abrupt yaw and roll to the right returned. The two pilots went through an emergency list, which included disengaging the yaw damper. The upset event ended soon afterward, and the airplane flew without further incident for the rest of the flight. ([10], p.52)

Examination of the maintenance records of the incident airplane revealed three rudder-related abnormal events during the month preceding the incident. The first event occurred on May 14 and was reported by the captain of the Eastwind incident flight himself. On that occasion, the captain experienced a series of uncommanded "taps" on the right rudder pedal just after takeoff, which he described as "like someone hitting their foot on the right rudder". The other two abnormal events happened on June 1 and June 8, respectively, and were reported by other pilots. They involved small, uncommanded yaw and roll of the airplane, although no movement of the rudder pedals was reported. The Eastwind incident happened on June 9, only a day after the latest maintenance inspection of the incident airplane's rudder control system. ([10], p.53)

When the NTSB examined Eastwind flight 517's rudder control system after the incident, they discovered a few anomalies related to the yaw damper component. First, the yaw damper had been rigged incorrectly. Instead of moving the rudder a maximum of 3 degrees in each direction, it could move the rudder only 1.5 degrees left but 4.5 degrees right. Second, the investigators found chafed wiring from the yaw damper to the main rudder PCU, which was consistent with but was not conclusive evidence of a short circuit. If there had been a short circuit, it could have caused an unintended yaw damper command to the rudder. ([18], p.209)

Could it be that an electrical fault in the yaw damper had caused the uncommanded yaw and roll of the Eastwind flight 517? The NTSB investigators thought it was unlikely. Even with the yaw damper misaligned, a yaw damper malfunction could only command a maximum of 4.5 degrees of rudder deflection to the right, which could not account for the difficulty experienced by the incident pilots in countering the roll and yaw. Representatives from Boeing, however, suggested that a yaw damper failure could be the cause of the incident and that the incident pilots were startled by the initial upset and exaggerated the difficulty in regaining control of the airplane. ([1], p.164)

To determine whether a yaw damper failure and an unintended yaw damper hardover (i.e., maximum yaw damper command) could account for the Eastwind flight 517 incident, the NTSB conducted flight tests using the incident airplane from June 22 to June 24, 1996. The main task of the flight tests was to document how the incident airplane and the incident pilots would respond to unexpected yaw damper hardover. For this purpose, the incident airplane's yaw damper remained misaligned, and it had a yaw damper fault-insertion device that allowed the FAA personnel to command a yaw damper hardover condition from the cockpit. ([10], p.61) During the tests, the captain of Eastwind flight 517 took control of the airplane and responded to a series of abrupt yaw damper hardover insertions made by the FAA personnel in the cockpit. ([10], p.62)

In each instance of yaw damper hardover, the captain of Eastwind flight 517 easily regained control of the airplane. ([10], p.62) When a Boeing representative asked him whether the yaw damper hardover resembled what he experienced during the incident, the captain responded: "This wasn't even close." ([1], p.165) The captain further indicated that during the incident, the rudder pedal felt stiffer and less effective, and the force required on the rudder pedal was much greater than during the flight tests. ([10], p.62)

If a yaw damper hardover could not produce the yaw and roll in the Eastwind flight 517 incident, what else could? When the investigators measured the Eastwind flight 517 rudder PCU servo valve, they discovered that it had relatively tight clearances, similar to the USAir flight 427 rudder PCU servo valve. ([10], p.264) Could it be that the two rudder PCU servo valves had both jammed via the same mechanism, i.e., thermal shock? The investigators hoped that the upcoming rudder PCU thermal testing would help to answer this question.

5.12 Rudder PCU Thermal Testing

In August and October 1996, the NTSB conducted two series of thermal tests to determine the effects of thermal shock on the operations of the dual concentric servo valve inside the main rudder PCU. The August tests were conducted at Canyon Engineering, a small hydraulic company in California. Due to Canyon's limited temperature control and data recording capabilities, the tests were repeated in October at Boeing's facility at Renton, using Boeing's superior equipment. ([1], p.172-173)

In both tests, the investigators first conducted the thermal tests on a newly produced PCU to verify setup and methodology, and then on the main rudder PCU from USAir flight 427. In each thermal test, the exterior temperature of the test PCU servo valve housing was cooled to -27 to -40 degrees F, which was estimated to be its temperature range at the time of the accident. The investigators then injected Hydraulic fluid into the test PCU, and they moved an input arm connected with the secondary (outer) slide of the test PCU servo valve to simulate a left or right rudder input command. ([10], p.78) There were three main test conditions, corresponding to three different temperature ranges of the hydraulic fluid:

1. The Baseline Test Condition: This test condition approximated the normal operating temperatures of the hydraulic fluid. Test data for this condition showed that the difference in the servo valve housing's exterior temperature and the hydraulic fluid temperature at the PCU inlet was approximately 50 to 60 degrees F. ([10], p.78)

2. The Simulated Hydraulic System Failure Condition: In this test condition, the temperature of the hydraulic fluid entering the PCU was raised to simulate a malfunction of one of the hydraulic pumps. The hydraulic fluid was first heated to 170 degrees F to simulate the effects of a pump failure. The fluid then passed through a 15 feet steel tube, which simulated the distance it needed to travel in the unheated tail fin before reaching the inlet on the rudder PCU. Test data for this condition showed that the

difference in the servo valve housing's exterior temperature and the hydraulic fluid temperature at the PCU inlet was approximately 100 degrees F. ([10], p.78)

3. The Extreme Temperature Differential Test Condition: This condition represented the "worst case" scenario and was not supposed to occur during normal flight operations. In this test condition, hydraulic fluid heated to 170 degrees F was injected directly into the inlet of a PCU at -40 degrees F. Test data showed that 25 seconds after the fluid insertion, the temperature difference between the hydraulic fluid and the servo valve housing could be as high as 180 degrees F. ([10], p.79)

In both the August and the October tests, the new rudder PCU behaved normally under all the three thermal test conditions. The main rudder PCU from USAir 427, in contrast, behaved normally under the first two test conditions but exhibited abnormal behaviors under the extreme temperature differential condition. In each extreme temperature differential test, The USAir 427 PCU was given a series of full left rudder commands. Its responses always followed the same pattern: The PCU would respond normally for the first few input commands. However, in the next few input cycles, the input arm that moved the secondary (outer) slide of the PCU servo valve would move slower than normal. At some point, the input arm would be stuck in the full left rudder position for a few seconds, and a significant amount of force (124 pounds in one case) was needed to return the input arm to its neutral position. ([10], p.79)

The thermal shock tests showed that the dual concentric servo valve within the main rudder PCU from USAir flight 427 could jam under the extreme temperature difference condition. More specifically, the extreme temperature difference could cause the secondary (outer) slide within the servo valve to jam to the servo valve housing.

After both the August and October test series, the NTSB investigators brought the USAir 427 main rudder PCU to Parker Hannifin's facility. They asked Parker's personnel to disassemble and examine the dual concentric servo valve. The investigators wanted to address the following question: Did the jamming of the secondary slide to the servo valve housing leave any witness marks on the servo valve components? This question was significant because when the investigators had first examined the servo valve after the flight 427 crash, they had not found any witness marks inside it. If they found witness marks now, it would show that (1) a thermal-shock-induced jam would leave witness marks and (2) there was no thermal-shock-induced jam on Flight 427 during the accident. ([1], p.173)

After examining the two slides and the interior of the servo valve housing, Parker technicians found no evidence of damage or physical marks from jamming during the thermal tests. Further, the servo valve slides still moved freely, and the servo valve was still capable of completing Parker Hannifin's functional acceptance test. ([10], p.245) This meant that the secondary slide from the accident PCU could jam to the servo valve housing without leaving any traces of physical evidence. ([10], p.246)

5.13 A New Rudder Reversal Mechanism

5.13.1 From Secondary Slide Jam to Rudder Reversal

Boeing was initially skeptical of the import of the thermal shock tests for a few reasons. First, the extreme temperature difference condition was unrealistic. It ignored the cooling effect on the heated hydraulic fluid as the fluid traveled from the hydraulic pumps to the rudder PCU. Second, there was no physical evidence that an extreme temperature difference needed for the jamming of the secondary slide existed in the accident airplane PCU at the initial upset. Finally, even if the secondary (outer) slide had jammed to the housing of the dual concentric servo valve in the accident PCU, the primary (inner) slide was expected to take over and compensate for the secondary slide's lack of movement . ([1], p.174) After all, the dual concentric servo valve design was supposed to provide redundancy in the unlikely event of the jamming of one of the two slides. ([18], p.120)

Despite their skepticism, Boeing engineers conducted a detailed examination of the data from the October 1996 thermal tests. The thermal tests had focused on the effects of thermal shock on the dual concentric servo valve, and Boeing engineers wanted to examine how the jamming of the servo valve's secondary slide would impact the movement of the rudder. However, a review of the hydraulic fluid flow data from the extreme temperature differential tests showed something alarming. When the secondary slide jammed to the servo valve housing and a further full rudder input was applied, the direction of the hydraulic flow within the rudder PCU corresponded to a rudder reversal (i.e., the rudder moves in the opposite direction to the pilot command) during the jam. ([10], p.245) If this was correct, it meant that a single jam within the servo valve could cause the main rudder PCU to fail catastrophically and that the PCU did not provide the redundancy that Boeing had promised. ([1], p.174)

Boeing engineers wanted to understand why the rudder reversed during the secondary slide jam. They modified a new-production PCU to simulate a secondary slide jam and then tested the effects of applying a full rudder input to this PCU. The tests showed that, when the secondary slide jammed to the servo valve housing at certain positions (specifically, more than 50% off the neutral position), a full rudder input would exceed the PCU's capability to respond and cause the internal input linkages to bend and twist. As a result, the primary slide would travel beyond its intended stop, to a position at which the PCU moved the rudder in the direction opposite of the intended command (i.e., reversal). ([10], p.81)

After knowing about Boeing's discoveries, the NTSB conducted its own rudder reversal tests in November 1996 on three rudder PCUs: a new-production PCU, the USAir flight 427 PCU, and the Eastwind flight 517 PCU. The investigators wanted to compare the three PCU's tendency to reversal *given* a secondary slide jam to determine whether the USAir flight 427 PCU was more susceptible to reversal than other servo valves. The tests showed that (1) all the three PCUs were capable of reversal given a secondary slide jam to the servo valve housing; (2) each PCU had a different "threshold of reversal", which was the minimum distance that the secondary slide had to be displaced from its neutral position[25] during the jam to result in a rudder reversal; (3) among the three PCUs, the USAir flight 427 PCU had the lowest threshold of reversal, and the thresholds for both the USAir flight 427 PCU and the Eastwind flight 517 PCU were significantly lower than the new-production PCU. ([10], p.84) In other words, USAir flight 427's PCU was indeed more vulnerable to rudder reversal than normal.

It was the second time the investigators discovered a rudder reversal mechanism in the Boeing 737 main rudder PCU. The first discovery was in 1992 when the main rudder PCU from the Mack Moore incident airplane was found capable of reversal. ([9], p.69) However, there are two main differences between the rudder reversal mechanisms found in these two cases. First, the reversal mechanism discovered in the Mack Moore incident involved the overtravel of the secondary slide. In contrast, the reversal mechanism discovered in the thermal shock tests involved the jamming of the secondary slide to the servo valve housing and the primary slide's overtravel.

Second, the rudder reversal mechanism in the Mack Moore incident was the result of a manufacturing defect unique to the incident airplane; neither the PCU from United flight 585 nor the PCU from USAir flight 427 was capable of reversal via the mechanism found in the Mack Moore incident. In contrast, the rudder reversal mechanism found in the thermal tests appeared to be the result of a design flaw: All the three PCUs examined by the NTSB were capable of reversal via this mechanism, even though the likelihood of reversal varied among the PCUs. Moreover, the existence of this rudder reversal mechanism meant that the PCU no longer protected against servo valve jams the way it was supposed to, which could mean that the Boeing 737 no longer met FAA's certification standards. For these reasons, Boeing took the discovery of the new rudder reversal mechanism very seriously, and immediately proposed short-term maintenance checks and long-term PCU redesigns to the FAA. ([10], p.222)

---

[25] The neutral position of the slides is the default position of the slides.

5.13.2 Boeing's Ground Demonstration

In June 1997, Boeing conducted a ground demonstration of the rudder jam and reversal, and all the parties in the USAir 427 investigation were invited to participate. The demonstration's purpose was to show what the secondary slide jam and the ensuing rudder reversal felt like to the pilots, and what the pilots could do to overcome the reversal. ([10], p.85) By this point, the Boeing engineers already knew that the rudder reversal's likelihood depended on where the secondary slide jammed to the servo valve housing. Hence, they fitted a newly manufactured 737-300 with a special tool to simulate a secondary slide jam at three different positions: About 0 percent, 25 percent, and 50 percent of a full travel from the neutral position. During the demonstration, each participant sat in the cockpit of the test airplane and manipulated the rudder pedals under the three simulated rudder jam conditions. ([10], p.86)

The first demonstration had the secondary slide jammed at 25 percent from its neutral position. When the participants pushed the rudder pedals slowly to their full positions, no reversal was produced, even though the left pedal was more difficult to push down than the right pedal. However, when the participants pushed the left rudder pedal hard or abruptly, the input triggered a rudder reversal most of the time. According to one participant from the NTSB, the left rudder pedal pushed back against his foot during the rudder reversal. "The motion was slightly slower than an input I would expect from a human." He noted. "The motion was steady and continued without pause no matter how hard I pushed to counter it. 'Unrelenting' was a description that, at the time, seemed to capture my impression." ([10], p.86) When the participants "stopped fighting" the rudder pedal's push back motion, the rudder reversal ended immediately, and the rudder pedals returned to their neutral position. ([10], p.86)

The second demonstration represented a secondary slide jam at its neutral position (i.e., 0 percent off). The participants noticed a slight difference between the two pedals (the right pedal slightly easier to push), although any resistance was easy to overcome. No matter how aggressively and abruptly the participants pressed the rudder pedals, no reversal was produced. ([10], p.86)

The third demonstration represented a secondary slide jam at 50 percent from its neutral position. Any abrupt motion on the pedals immediately initiated a rudder reversal. Even when the participants pressed the pedals slowly and steadily, sometimes a rudder reversal still occurred. Moreover, the participants discovered that the rudder reversal was faster than with a jam at the 25 percent position. According to one participant, “It was impossible to stop the motion by physically pushing against the rudder pedal. One several trials, I tried relaxing my input momentarily before the rudder pedal reached the upper stop. I found the rudder reversal motion continued. This was not true in the jam at the 25 percent position, when the relaxation of the pressure seemed to automatically stop the reversal motion. This motion was faster, easier to initiate, and more difficult to stop.” ([10], p.87)

Another participant stated that, during the third demonstration (that represented a secondary slide jam at about 50 percent off the neutral position), he experimented with switching the hydraulic system to “standby”, allowing the standby rudder PCU to take control of the rudder over the jammed main rudder PCU. He discovered that this action eliminated the rudder reversal, allowing him to recenter the rudder by rudder pedal inputs in the normal direction. During subsequent rudder movements with the standby rudder system engaged, the rudder did not reverse. ([10], p.87)

The Boeing ground demonstrations confirmed that rudder reversals became more severe the farther the secondary slide jammed off its neutral position. Moreover, when the secondary slide jammed at 50 percent of full travel or more from the neutral position, it became physically impossible for the pilots to resist the reversal motion on the rudder pedals, and the only reliable way to stop the reversal was to immediately switch from the jammed main rudder PCU to the standby rudder PCU. Unless the pilots understood what was happening, the most severe rudder reversals could not be stopped.

5.14 Two Competing Scenarios

In the fall of 1996, after the rudder PCU thermal tests and the subsequent discovery of a rudder reversal mechanism, the NTSB investigators arrived at the following picture of what had caused the crash of USAir flight 427: ([18], p.231)

1. Flight 427 encountered Delta flight 1083's wake vortices and rolled left as a result.
2. The pilots applied the right rudder pedal to correct the roll.
3. The secondary slide within the main rudder PCU servo valve jammed to the servo valve housing, likely at more than 50 percent of a full travel from its neutral position.
4. The combination of the jam and the pilot input led to a rudder reversal.
5. The rudder reversal was a hardover—a sustained deflection of the rudder at full left position.
6. The hardover happened at about 190 knots, close to the accident aircraft's crossover speed, below which the control wheel could not overcome the effect of rudder hardover.
7. The aircraft yawed and rolled left drastically, stalled, lost control, and crashed.

Boeing, however, disagreed with the NTSB's scenario. The thermal tests and the discovery of the second rudder reversal mechanism did persuade Boeing that its 737 main rudder PCU had a problem. However, Boeing still resisted the conclusion that a jam and reversal of the main rudder PCU had played a role in the USAir flight 427 crash. Instead, it argued as follows: ([18], p.231-232)

1. The thermal tests showed that the main rudder PCU from USAir flight 427 could only jam under the extreme thermal differential condition.
2. The extreme thermal differential condition was unrealistic and highly unlikely to be encountered in normal flight operations.

3. There was no evidence that the rudder PCU from USAir flight 427 had jammed or reversed.
4. A rudder reversal due to a secondary slide jam had never been documented in the history of the Boeing 737.
5. The pilots, startled by the sudden left roll caused by the wake vortices from Delta flight 1083, might have pressed the wrong (i.e., left) rudder pedal and kept it pressed during the accident.
6. There is evidence from the psychological literature that startled people could make mistakes.

The disagreement between the NTSB and Boeing had reached an impasse. Both the NTSB's scenario and Boeing's scenario postulated hypothetical and improbable events, but neither side provided evidence for the actual occurrences of these hypothetical events. On the one hand, the NTSB's scenario assumed that the secondary slide within the main rudder PCU servo valve had jammed to the servo valve housing. The thermal tests had shown that (1) the PCU *could have* jammed in this way if there had been an extreme thermal shock condition, (2) the PCU *could have* jammed without leaving any physical traces behind. Nevertheless, Boeing was correct in pointing out that the NTSB did not have evidence that the PCU had jammed this way, or that the extreme thermal shock condition had obtained. ([1], p.174)

On the other hand, Boeing's scenario assumed that the pilots had been startled (by the wake turbulence) and applied the wrong rudder pedals for an extended amount of time. However, Boeing had no evidence for this assumption, either. To explain why the pilots would continue to press the wrong rudder pedal as the yaw and roll continued, Boeing proposed various possible explanations, including pilot incapacitation (due to medical conditions), spatial disorientation, deliberate pilot action, and unintended rudder pedal activation. However, the examination of the pilots' background and the circumstances of the accident found all these possible explanations highly improbable. ([10], p.253-255)

It would have been easy to determine which scenario was correct if USAir flight 427's FDR had recorded the pilot's rudder inputs: The investigators could have checked the FDR data to see whether the pilots had applied the left or the right rudder pedal. Similarly, it would have helped if the accident airplane's FDR had recorded the time histories of rudder positions, since pilot-induced rudder movements differed in characteristics from rudder reversals. Unfortunately, the FDR from USAir flight 427 had recorded neither the pilot rudder inputs nor the time histories of rudder positions. To compensate for this limitation in the FDR data, the NTSB and Boeing attempted to derive the rudder positions and pilot rudder inputs from the parameters that the FDR did measure (e.g., heading, pitch, and roll.), using techniques including kinematics and computer workstation simulations. ([10], p.91)

5.15 Kinematics and Computer Simulations

During the USAir flight 427 investigation, Boeing applied the kinematics analysis technique to derive from available FDR data the position of flight control surfaces that were not recorded by the FDR. Boeing's kinematics process involved fitting curves through available FDR data such as heading, roll, and pitch, obtaining time histories of the rates from these curves; and deriving accelerations from these rates. These accelerations allowed Boeing to calculate forces, moments and aerodynamic coefficients using Newton's laws of physics, and finally to derive flight control time histories using its aerodynamic models. ([10], p.87)

On September 30, 1997, Boeing formally submitted its kinematic solution for the USAir flight 427 accident to the NTSB. Boeing's scenario assumed that about 19:02:58, the flight crew applied considerable right control wheel because of the left roll produced by the wake vortex encounter. The right control wheel arrested the left roll and initiated a right roll. Boeing's scenario then assumed that the flight crew applied the left rudder pedal about 19:02:59, resulting in a left rudder movement of about 12 degrees just before 19:03:00. About 19:03:00, the flight crew temporarily released the control wheel and the left pedal inputs, and the left rudder deflection was reduced to about 3 degrees. However, the airplane was still in the wake vortices, which rolled the airplane to the left. To counter the left roll, the first officer again applied the right control wheel; however, the airplane continued to accelerate in a left roll. Finally, Boeing's scenario assumed that, between 19:03:00 and 19:03:01, the flight crew again applied full left rudder pedal pressure and maintained the left rudder pedal input until ground contact. ([10], p.92)

Because Boeing was one of the few entities in the world that possessed the technological capacity and knowledge of the 737 airplanes to reconstruct details of potential accident scenarios, the NTSB took the kinematic solution submitted by Boeing seriously. ([10], p.90) Nevertheless, the NTSB noted a few weaknesses of Boeing's kinematic analysis. First, review of Boeing's kinematic process showed that, due to the short term duration of the flight 427 crash and the relatively infrequent sampling of the FDR heading data (recorded once-per-second), there were not enough heading samples to perform the kinematic calculations effectively. To solve this problem, Boeing used interpolation techniques to curve fit the FDR heading data, thereby creating more artificial data between the FDR recorded data points. The use of interpolation techniques introduced potential errors since different interpolation techniques could result in different rudder surface time histories. Second, kinematic analysis magnified the noise inherent in the FDR data, and Boeing had to use its data smoothing program to reduce the noise in the data. ([10], p.88)

Because of these weaknesses of the kinematics approach, the NTSB chose to use a different method to reconstruct the flight control time histories of USAir flight 427. The investigators tried out different flight control surface histories in the flight simulation software that ran on the NTSB's computer workstations, to find the flight control time histories that provided the best match with the FDR data. The method is formally known as iteration and works as follows. The NTSB investigators used some values of flight control positions as initial inputs into its computer simulations. They then compared the output of the simulations—such as heading, airspeed, and vertical acceleration—with the available FDR data. They then modified the control input, re-ran the simulations, and continued this process until it obtained a good match with the FDR data. ([10], p.88)

Using iteration, the NTSB found a computer simulation scenario (subsequently referred to as the NTSB's best match simulation) with an excellent match with the USAir flight 427 FDR data. According to the NTSB's best match simulation, just after 19:02:58, the wake vortex produced a left yaw motion. The simulation assumed that the flight crew responded to the left yaw motion with a right rudder pedal input at about 19:03:00. This scenario further assumed that the secondary slide jammed to the servo valve housing at the 100 percent travel position, and the flight crews' rudder pedal input resulted in a rudder reversal. The rudder reached its left blowdown limit at about 19:03:00 and remained at that position until 19:03:08 when the airplane stalled. ([10], p.91-92)

However, the NTSB could not be sure that the control surface positions derived from its computer simulations reflected the actual control surface positions, for two reasons. First, even though the NTSB's computer simulations did not use interpolation or data smoothing techniques, there were still various factors that limited the extent to which the simulations accurately reflect the accident flight history. For example, the computer simulation software encoded various aerodynamic models of the airplane. The aerodynamic models had been validated to some degree by the flight tests, but safety factors had limited the validation process. For another example, even though the computer simulations had taken into account wake turbulence, the respective contributions of the flight control surfaces and wake turbulence remained uncertain. ([10], p.88)

Second, when the NTSB investigators compared its best-match simulation with Boeing's kinematic solution, they found that both solutions matched the FDR data equally well.

([10], p.246) Since the FDR data could not decide between the two scenarios, another source of evidence was needed to determine which was correct.

5.16 Speech and Breathing Analysis

Surprisingly, the evidence that ended up deciding between the NTSB scenario and the Boeing scenario came from the Human Performance Group. As part of its investigation, the Human Performance Group analyzed the pilot speech (including pitch, amplitude, and speaking rate) and breathing (inhaling, exhaling and grunting) patterns recorded by the CVR, in order to understand the behaviors and psychological states of the pilots during the accident sequence. ([10], p.136) The Human Performance Group investigators then consulted three independent specialists to validate its speech and breathing analysis. ([10], p.140) Finally, the investigators evaluated how well the rudder and control wheel time histories produced by the NTSB computer simulations and Boeing's kinematic solution fit the speech and breathing analysis' results. ([10], p.246)

First, to evaluate Boeing's proposal that one of the flight crew panicked or became "over-aroused" because of the wake turbulence encounter and pressed on the wrong rudder pedal, the Human Performance Group analyzed the CVR tape and measured the fundamental frequency (pitch) and amplitude (volume) of each sentence. According to scientific literature, fundamental frequency and amplitude tend to increase in response to increased psychological stress. ([10], p.136) If the flight crew had panicked during the wake turbulence encounter, there should be a sharp increase in their utterances' fundamental frequency and amplitude at the initial upset.

A review of the CVR tape showed that the first officer did not speak enough during the emergency period for a meaningful analysis, so the speech analysis focused on the captain's statements. The investigators compared the fundamental frequencies of the captain's statements before and during the initial upset. They found an increase in fundamental frequency at the initial upset, from about 140 Hz to about 200 Hz. The magnitude of the increase indicated surprise and increased stress, but not at the level of panic. Moreover, the investigators saw a gradual increase in the fundamental frequency after the initial upset until the last few seconds before the crash, when the captain was screaming, and the fundamental frequency of his speech exceeded 350 Hz. ([10], p.136) These discoveries indicated that the captain's stress level had progressively increased after the initial upset and that he had not become "over-aroused" at the initial upset. ([18], p.184)

Second, even though the first officer (the pilot flying) spoke little during the accident sequence, his breathing patterns—in particular, the grunting and forced exhalations—were highly informative. Between about 19:03:00 and about 19:03:02, the CVR recorded the sounds of grunting on the first officer's hot microphone channel. Two specialists in breathing physiology who examined the grunts stated that they were signs of significant physical effort, much greater than sounds produced by normal use of flight controls. One specialist concluded that the sound suggested that the first officer was struggling unusually hard as if he was experiencing strong resistance in flight control. ([10], p.250)

When the Human Performance investigators put the first officer's breathing and grunts on a timeline, they found that the grunts' timeline did not match Boeing's scenario at all. According to Boeing's scenario, the grunting sounds occurred after the first officer made a full right control wheel input and after a hypothesized left rudder input. However, neither of the two maneuvers would have required more than 70 pounds of force, which was relatively mild and should not cause a pilot to grunt. A pilot could exert more than 70 pounds of force in holding a rudder at full deflection, but there would be no reason to do so if the rudder was functioning normally and a full left rudder deflection had occurred. The investigators were unable to explain the first officer's loud grunting sounds given Boeing's scenario. ([10], p.255)

In contrast, the timeline of the first officer's grunting sounds perfectly matched the NTSB's rudder reversal scenario. The CVR recorded the first grunting sound at 19:03:00.3, about 0.4 seconds after the start of the rudder reversal postulated by the NTSB computer simulation. The grunt was relatively soft and could manifest an involuntary physical reaction by the first officer to the start of the rudder pedal pushing back against his right foot. ([10], p.250)

Moreover, the CVR recorded louder grunting sounds by the first officer at 19:03:01.5, about 0.6 seconds after the rudder reversal had reached hardover. The right rudder pedal had fully pushed back with maximum displacement. According to the NTSB's analysis, the push back force on the rudder pedal at this time could reach 400 pounds, which would explain why the first officer was grunting so loudly: He was exerting an immense effort to resist the push back pressure on the rudder pedal with no apparent effect. ([10], p.251)

The Human Performance Group's analysis of the pilots' speech and breathing patterns fit well with the NTSB's rudder reversal scenario but was inconsistent with Boeing's incorrect pilot input scenario. It was the NTSB's most persuasive evidence that the rudder had reversed in the USAir flight 427 accident.

5.17 Resolving United 585 and Eastwind 517

After its breakthrough in the USAir flight 427 investigation, the NTSB further conducted computer simulation and human performance analysis of the United flight 585 crash and the Eastwind 517 incident. The three cases' flight profiles bore strong similarities to each other, and the investigators wanted to determine whether rudder reversal was a common cause.

5.17.1 United Flight 585

The NTSB conducted computer simulations to reconstruct United flight 585's flight control input histories based on the FDR and radar data. The FDR on United flight 585 recorded even fewer parameters than the one on USAir flight 427, making the reconstruction more difficult. Further, United flight 585 likely encountered mountain-generated turbulence and gust, but the exact wind characteristics during the encounter were unknown. ([10], p.258) Despite these limitations, The NTSB was able to come up with a computer simulation of a rudder reversal scenario that had an excellent match with the FDR data. For instance, the heading output from the best-match simulation matched the FDR heading data within 1 degree or less throughout the accident sequence. ([10], p.259)

According to the FDR, about 09:43:30 (11 seconds before ground impact), the accident airplane's heading began to move right and continued at 4.7 degrees per second for 3 seconds. The NTSB's best-match computer simulation assumed that a strong crosswind caused this right yaw and that the captain (the pilot flying) responded to the heading change by applying left rudder pedal input about 09:43:32. The input's timing was consistent with the time needed for the captain to perceive the yaw, wait a moment for the effect of the turbulence to subside, and decide that a left rudder input was required. The best-match simulation further assumed that the secondary slide jammed at 100 percent off its neutral position, and that the captain's left rudder input initiated a rudder reversal to the right, which reached hardover at about 09:43:33. ([10], p.261)

The CVR tape on United flight 585 had lower audio quality compared to USAir flight 427. It did not pick up the pilots' breathing patterns, which prevented the Human Performance Group from determining whether the captain had made any grunting sounds at the moment of rudder reversal. However, the CVR data indicated that at 09:43:33.5 (about 1.5 seconds after the rudder reversal began), the captain said "fifteen flaps", which signaled his decision to abort the landing. At this moment, the airplane's bank angle had not exceeded 20 degrees, and the pitch was 8 degrees nose down, still within the range of normal descent. However, speech analysis of the captain's statement showed a heightened fundamental frequency that was consistent with a sense of urgency. The sense of urgency was further suggested by the captain's omission of a call-out item in the normal go-around procedure, which could be the result of the captain's struggle with the rudder reversal. ([10], p.261)

Because of the limited FDR data, the NTSB was able to identify two alternative scenarios that matched the FDR data as well as the rudder reversal scenario. One scenario assumed that the right control wheel alone (without rudder input) caused the right roll to worsen after 09:43:32. However, this scenario was not realistic, because it assumed that the flight crew flew the airplane into the ground when simple wheel corrections could prevent the right roll. Examination of the CVR tape showed no evidence that the pilots of United flight 585 deliberately crashed the airplane. ([10], p.259)

The other scenario was submitted by Boeing in June 1997 and assumed that the continued right roll was caused by a rotor that followed the flight path of the accident airplane for 8 seconds and increased in strength to about 1.8 radians (103 degrees) per second. ([10], p.259) However, the NTSB pointed out that the rotor postulated by Boeing's scenario was exceptionally severe: The strongest rotors ever documented in the Colorado Springs area had a strength of about 0.05 radians per second. Moreover, if United flight 585 had penetrated the rotor's low-pressure core and remained there for 8 seconds, its FDR would have recorded signature changes in indicated airspeed and altitude. Its CVR would have recorded sounds characteristic of intense rotors. None of these traces could be found in the FDR and CVR data from the accident airplane. ([10], p.260)

Based on these considerations, the NTSB concluded that a rudder reversal had caused the United flight 585 crash. ([10], p.263)

5.17.2 Eastwind Flight 517

Unlike United flight 585 and USAir flight 427, Eastwind flight 517 did not crash, and the flight crew was able to regain control of the airplane shortly after the initial upset. This was partly because Eastwind flight 517 was flying at a speed well above its crossover airspeed when the incident occurred, and the control wheel had sufficient roll control authority to overcome the effects of a rudder deflection. ([10], p.269) Nevertheless, Eastwind flight 517's FDR data and its flight crew's statements strongly suggested that a rudder reversal had occurred. So the NTSB used iteration to find a computer simulation of the rudder reversal scenario that best matched the FDR data. ([10], p.116)

According to the NTSB's best-match simulation, the upset was initiated by an uncommanded yaw damper hardover that caused a 3.95 degree right rudder deflection. Two seconds later, the pilots stepped on the left rudder pedal to the counter the uncommanded right yaw. This scenario further assumed that the secondary slide was jammed to the servo valve housing about 55 percent off neutral. When the pilot applied force to the left pedal, the rudder reversed and deflected to its blowdown limit to the right. Due to a hydraulic fluid leakage problem found in the main rudder PCU, the blowdown limit was reduced to 6.5 degrees of rudder deflection, which likely made the uncommanded right yaw less severe than it would have been. ([10], p.265)

The NTSB's simulation result closely matched the FDR's heading, roll, and vertical acceleration data. ([10], p.265) Moreover, the rudder pedal force assumed in the computer simulation was consistent with the flight crew's post-incident statements. In post-incident interviews with the NTSB investigators, the captain of Eastwind flight 517 stated that he immediately applied "opposite [left] rudder and stood pretty hard on the pedal" after the upset. He said that the "rudder moved, but felt stiffer than normal." According to the first officer, the captain was "fighting, trying to regain control" and "standing on the left rudder [pedal]." ([10], p.267) The NTSB's best-match simulation assumed that the left rudder pedal force increased to 500 pounds after the rudder reversal, consistent with both pilots' reports that the captain had to exert substantial force on the left rudder pedal by "standing on" it. ([10], p.268)

In a kinematic solution submitted to the NTSB in August 1998, Boeing proposed an alternative scenario of the Eastwind flight 517 incident. ([10], p.265) According to Boeing's kinematics solution, the captain of Eastwind flight 517 responded to a yaw damper related left yaw by applying the right rudder pedal, and then maintained a light right rudder pedal pressure (about 54 pounds) for over 10 seconds. Boeing's kinematics solution matched the

FDR data as well as the NTSB simulation; however, the NTSB pointed out that Boeing's scenario contained numerous inconsistencies with the human performance data. Most importantly, it was highly unlikely that the flight crew would report applying left rudder when they had applied right rudder, and even less likely that they would recall "standing on the left rudder" when the captain had applied only a small force on the right rudder pedal. ([10], p.271)

Therefore, based on its computer simulation and human performance analysis, the NTSB concluded that the rudder had reversed during the Eastwind 517 incident. ([10], p.271)

5.18 Conclusion

On March 24, 1999, about four and a half years after the accident, the NTSB issued its final report for the USAir flight 427 investigation. ([18], p.261) The "probable cause" statement of the accident report reads as follows:

> The National Transportation Safety Board determines that the probable cause of the USAir flight 427 accident was a loss of control of the airplane resulting from the movement of the rudder surface to its blowdown limit. The rudder surface most likely deflected in a direction opposite to that commanded by the pilots as a result of a jam of the main rudder power control unit servo valve secondary slide to the servo valve housing offset from its neutral position and overtravel of the primary slide. ([10], p.295)

In the report, the NTSB made ten safety recommendations to the FAA. First, the NTSB argued that the dual concentric servo valve used in all the Boeing 737 main rudder PCUs is not "reliably redundant". The first recommendation was to require that all the current and future 737s have a reliably redundant rudder system. Second, the NTSB recommended that the FAA convene an engineering test and evaluation board (ETEB) to conduct further analysis of the Boeing 737 rudder system's potential failure modes. The third and fourth recommendations were aimed at passenger aircraft generally. They suggested that they all have a reliably redundant rudder system and the capacity to continue safe flight in the event of any jammed control surface. ([18], p.262)

The NTSB report exonerated the pilots of USAir flight 427 (and the pilots of United flight 585), arguing that they had responded to the upset promptly, and that they could not be expected to correctly diagnose the rudder control problem and then recover from the rudder reversal given the circumstance of the accident. ([10], p.293) However, the report further argued that the training provided to Boeing 737 pilots and the existing recovery procedure from a jammed rudder were inadequate, and that the pilots did not have adequate knowledge about the crossover airspeed hazard. As a result, the NTSB made two recommendations aiming at improving the training of the pilots to handle rudder jams, plus two recommendations to ensure that the pilots would maintain safety margins above the crossover speed in flight. ([10], p.297)

Finally, the report criticized the FAA's failure to require timely and aggressive actions to increase the flight data recorder capacity. It argued that the lack of recorded parameters had significantly hampered the NTSB's ability to identify safety hazards the United flight 585 and USAir flight 427 accidents. ([10], p.295) Consequently, the final two recommendations were about drastically increasing the number of flight parameters measured by the flight data recorders on all US airlines by the fall of 2001. ([10], p.297)

In May 1999, two months after the NTSB's recommendations, the FAA formed its Engineering Test and Evaluation Board (ETEB) to investigate further the potential hazards of the 737's rudder system. On July 20, 2000, the ETEB submitted a 950-page report to the FAA. The report listed 46 failure modes and jam mechanisms in the 737 rudder system that could have catastrophic effects, especially during landing and takeoff. Many of these vulnerabilities were discovered for the first time and went beyond the failure mechanisms identified in the NTSB investigations. Echoing the NTSB report for USAir flight 427, the ETEB report suggested that the existing maintenance procedures could not catch some of the failures, that the pilots were poorly trained to deal with them, and that the flight data recorders did not record enough parameters to monitor critical system functions. ([18], p.269) On September 13, 2000, the FAA announced that it would require Boeing to redesign the rudder control system for all models of the 737. It would also implement the new pilot training programs, operational procedures, and maintenance checks proposed by the NTSB and ETEB recommendations. ([18], p.275)

Finally, on March 21, 2001, 10 years after the United flight 585 accident, the NTSB issued a revised report, concluding that the accident's probable cause was a rudder reversal. ([9], p.139) The case that had been unresolved for over ten years had finally come to a close.

Chapter 6

Case Study 5: TWA Flight 800

6.1 History of The Flight

On July 17, 1996, at 20:31 eastern daylight time (EDT), Trans World Airlines Flight 800, a Boeing 747-100, exploded in the air and crashed in the Atlantic Ocean near East Moriches, New York. The accident was one of the worst in-flight breakups in the U. S aviation history. The four-year investigation by the National Transportation Safety Board (NTSB) concluded on August 20, 2000, ending the most prolonged and complex air disaster investigation in U. S history to that time. ([6], p.59)

On the day of the accident, the accident airplane flew from Athens, Greece, and arrived at John. F. Kennedy International Airport (JFK) at 16:38 EDT. The temperature on the ground exceeded 80F. The airplane's next flight as TWA flight 800 was scheduled to take off at 19:00. To keep the cabin cool, aircraft operators ran two of the air-conditioning packs for about 2 1/2 hours. ([3] p.1)

TWA 800 was scheduled to depart for Paris, France. The flight did not require refueling, so there is only a negligible quantity of fuel left from the previous flight in the center wing fuel tank (CWT). The flight was to depart at 19:00, but it was delayed because a disabled piece of ground equipment blocked the airplane at the gate, and because of concerns about a suspected passenger/baggage mismatch. At 20:19, TWA 800 departed JFK, with 230 people on board. ([3] p.1)

The aircraft climbed to its assigned altitude of 13,000 feet uneventfully. But at 20:29, the Cockpit Voice Recorder (CVR) recorded the captain saying: "Look at that crazy fuel flow indicator there on number four ... see that?" Soon, the flight crew received instruction from Boston Air Traffic Control Center to climb to 15,000 feet. As the 747 approached 14,000 feet, the CVR recording of the next 30 seconds from the cockpit area microphone (CAM) includes the following sounds: A sound similar to a mechanical movement in the cockpit (at 20:30:42), an unintelligible word (at 20:31:03), and sounds resembling recording tape damaging noise (at 20:31:05). The CVR and FDR recordings then ended abruptly at 20:31:12. ([13], p.1-3)

At the time, TWA flight 800 had been flying in clear weather over the ocean near East Moriches, New York. Radar data and ATC (Air Traffic Control) showed that an Eastwind Airlines Boeing-737 (Stinger Bee Flight 507) was about 20 to 25 miles northeast of TWA flight 800, heading in the southwesterly direction. According to the Boston ARTCC (Air Route Traffic) transcript, at 20:31:50, the captain of Stinger Bee flight 507 reported that he "just saw an explosion out here." About 10 seconds later, the captain further reported, "we just saw an explosion up ahead of us here...about 16,000 feet or something like that, it just went down into the water". Later, many other pilots operating in the area reported an explosion to the air traffic control facilities in the area. ([13], p.3)

Many witnesses near the accident site reported seeing explosions, a large fireball in the sky, and debris (some of which burning) falling into the water. According to the FBI witness interviews, about one-third of these witnesses reported a streak of light moving upward in the sky to where a large fireball appeared. Several witnesses reported seeing this fireball split into two fireballs as it descended towards the water. ([13], p.3) Together with the widespread distribution of wreckage with a four-mile radius beneath the surface of the Atlantic Ocean, these witness interviews were the first indications that TWA 800 had experienced a catastrophic in-flight breakup. The airplane had been airborne for only 12 minutes. ([13], p.1)

6.2 Initial Investigation

6.2.1 Search and Recovery

The National Transportation Safety Board (NTSB), a U.S. government investigative agency responsible for transportation accident investigation, was notified about 20:50 the day of the accident. A full go team assembled and arrived on the scene early the next morning. ([13], p.313) Meanwhile, early reports of explosion led many to believe that the crash was caused by a bomb or a missile attack. As a result, the FBI started a parallel criminal investigation alongside the NTSB accident investigation. ([6], p.59)

Various agencies conducted search and recovery operations. Various civilian and military vessels reached the crash site and searched for survivors within minutes of the initial water impact, but found none. After the recovery of victims' bodies by scuba divers and remote-operated vehicles (ROVs), the U. S Navy spent over nine months recovering over 20,000 pieces of wreckage, approximately 95% of the total plane. Ships with sonar began by mapping the distribution of metal objects on the ocean floor. Divers and ROVs then systematically removed all visible debris from a water depth of 120 feet. After that, scallop dredges—steel sleds trailed by nets—scoured 40 square miles of the ocean floor for almost six months until they found no new materials. ([6], p.68-69.) The search and recovery identified three major areas of underwater wreckage: The yellow zone, red zone, and green zone corresponding to wreckage from the front, center, and rear sections of the airplane, respectively. ([13], p.65) The red zone is closest to JFK airport along the airplane's flight path, whereas the green zone is located furthest from JFK along the flight path. ([13], p.71-74)

After the wreckage recovery, the next challenge was to identify the parts. The NTSB investigators brought the wreckage to an abandoned hangar complex at Calverton, Long Island, where it was spread out on the floor and carefully examined. Using a Boeing computer database for identification, they would typically scroll through computer drawings to identify a particular part. Sometimes, a part had a serial number or a station number, which showed its location relative to the airplane's nose. Other hints, such as paint finish and sheet metal thickness, can also guide the jigsaw puzzle of part identification. For instance, the paint on top of the aircraft is more exposed to the weather and is duller than the paint on bottom surfaces. Thicker parts have greater load-bearing capacity and are more likely to have specific locations on the fuselage, etc. ([6], p.69) Once they were identified, the pieces were laid out on a two-dimensional grid on the hangar floor.

Also, to help determine the sequence of events of the failure, the investigators performed three-dimensional reconstruction of parts of the fuselage. In the largest hanger within the complex, a contractor fabricated a steel structure to hang the fragments; the framework resembled the ribs of a gigantic animal and was called "Jetosaurus Rex". After the recovery and identification of the wreckage, it took an additional three months to reconstruct a 94 feet section of the accident airplane's fuselage and some of its internal structures. ([6], p.69) A second hanger was used to reconstruct the aircraft cabin, including galleys, bathrooms, and passenger seats. ([28], Chapter 17) This second reconstruction helped determine whether there was any evidence in the cabin area showing a high-energy explosion. ([14])

6.2.2 Witness Interviews

In the hours that followed the accident, many people had witnessed the crash of TWA Flight 800. Even though there are discrepancies between different accounts, many witnesses saw a "streak of light" moving upward in the sky until a large fireball appeared. The public was intensely interested in these witness reports and speculated that the streak of light was a missile that struck TWA flight 800 and blew it apart. These witness accounts were a major cause of the initiation of the FBI criminal investigation. The FBI began interviewing witnesses on the evening of the accident and, within a week, had contacted over 500 witnesses. ([13], p.262)

The FBI conducted the witness interviews to determine if a missile struck TWA 800. The FBI agents did not produce verbatim records of the witness interviews. Instead, those who conducted the interviews wrote summaries, and they did not ask the witnesses to review or correct them. Some suggested interview questions provided by the FBI to its agents include assumptions about the missile attack, for example, "How long did the missile fly?" and "What does the terrain around the launch site look like?" ([13], p.266)

Even though NTSB expressed intent to form its witness group and interview witnesses, the FBI raised multiple concerns, including the non-government parties in the NTSB investigation having access to the witness information, and legal complications from multiple interviews of the same witness. The NTSB deferred, and the FBI conducted its witness interviews without including NTSB investigators. In November 1996, the FBI allowed the NTSB access to witness accounts' summaries with identifying information redacted. In April 1998, the FBI gave the NTSB the identities of the witnesses. However, because of the time elapsed, NTSB relied on the original FBI documents rather than re-interview witnesses. ([13], p.266)

6.3 Reconstruction of the Failure Sequence

To address how the airplane broke up, NTSB formed the Structures and Sequencing Group, with members from the major parties to the investigation. Representatives from the FBI also monitored but did not take part in the sequencing group's work. The sequencing group was to find out how the airplane broke apart and where the breakup started so that investigation efforts could concentrate on the causes of the breakup. ([14]) The Sequencing Group did most of its examinations on the wreckage as it was placed on the two-dimensional grid and the partial three-dimensional reconstructions.

Initial examination of the reconstructed portion of the airplane showed a relatively clear demarcation between pieces in red, yellow, and green zones. The recovery locations of the three zones relative to the flight path showed: First, the red zone pieces (including the forward portion of the wing center section (WCS) and a portion of fuselage directly in front of the WCS) were the earliest pieces to separate from the airplane. Next, the yellow zone pieces (the fuselage forward the wing center section) departed shortly after the red zone piece. Finally, the green zone pieces (wings and the aft portion of the fuselage) remained intact for some time after the forward fuselage's separation and impacted the water in the green zone. ([13], p.69-73) Since the area in and around the wing center section (WCS) was the first to depart the airplane, the breakup must have initiated in this area. ([13], p.260)

The 747-100's wing center section (WCS) is a large box structure about the size of a two-car garage. It extends between two walls called the front spar and the rear spar, and the side-of-body ribs that separate it from the wing fuel tanks. Its upper skin panels separate the WCS from the passenger cabin floor (which is above WCS and supported by floor beams). The lower skin panels separate the WCS from the airplane's air conditioning equipment (which is below the WCS). Internally, the wing center section divides into compartments by a series of lateral, or spanwise, beams. The center wing fuel tank (CWT) occupies most of the wing center section. It extends from the rear spar to spanwise beam 3, which is just behind the front spar. ([13], p.12-16) Finally, two large storage bottles for drinking water carried on the airplane were located right at the forward side of the front spar.

To determine how the wing center section (WCS) failed, the sequencing group began by examining portions of the structure in great detail. The investigators then developed localized sequences of events based on the observed features in each portion. Eventually, they combined individual local sequences of events until a coherent overall breakup sequence emerged. Using stress analysis, the investigators made sure that the proposed breakup sequence was consistent with the materials' structural properties and expected failure modes. ([13], p.103)

Based on a detailed examination of the WCS structure and supported by the results of stress analysis, the Sequencing Group concluded that an overpressure event initiated the breakup of the airplane. First, the investigators determined that spanwise beam 3 fractured at the top and rotated forward. Spanwise beam 3 is the most forward and weakest boundary member of the center wing tank, and it sits right behind the front spar of the wing center section. ([13], p.103) The forward rotation of the spanwise beam 3 was evident because the line of rivet heads running along the top of the beam left a series of "witness marks" on the aft side of the front spar. The witness marks comprised a series of evenly spaced, penny-sized depressions that exactly matched the line of rivet heads. The witness marks extended across nearly the entire the front spar, implying that the entire spanwise beam 3 was rotating forward. These features are evidence of excessive pressure on the aft side of the spanwise beam 3. ([13], p.106)

Second, the investigators found that the front spar itself bulged forward in two lobes, restricted by the inertial resistance of the water storage bottles mounted on its front side. The bulging and its shape indicated the escaping overpressure within the center wing tank on the aft side of the front spar. ([13], 103-106) Finally, the upper skin panel of the wing center section bulged upward as spanwise beam 3 separated. Evidence for this upward bulging included the pattern of the impact mark created by the upper end of spanwise beam 3 when it struck the stiffener immediately forward of the beam's upper end. This upward lifting of the skin again showed excessive pressure within the center wing tank. In short, the evidence led to the conclusion that an overpressure event within the center wing tank caused the fracture of the spanwise beam 3 at its upper end, and the separated beam rotated forward and contacted the front spar. ([13], p.106)

The investigators further traced the direction of cracking along these early fractures, by examining the features of the rivet-to-rivet fracture pattern of the fuselage skin in these areas. For instance, if the pieces on either side of the fracture are bending or deforming next to each other, then there must be other fractures somewhere else, causing these pieces to bend and fracture. In contrast, fractures that occur early are distinctly different, in that pieces on either side of the breaks pulled apart either in direct tension with no deformation or else in an unzipping effect. Such distinguishing features allowed the investigators to characterize the timing of the fractures in the failure sequence. ([14]) The sequence of events determined by the investigators is:

After the impact from spanwise beam 3, the front spar fractured at its upper end, and the fractures progressed down the front spar and spread to portions of the fuselage skin forward of the front spar. After cracking started in the fuselage at this location, it quickly spread through the lower fuselage. These early fractures created a large hole in the airplane's belly, through which structure and interior components (including the front spar and spanwise beam 3) ejected from the airplane and landed in the red zone. This entire portion of the sequence occurred quickly, within only a few seconds. After the belly structure's loss, fractures progressed up the sides and across the top of the fuselage. Finally, the nose portion of the airplane separated and fell into the yellow zone. ([13], p.108-109)

The airplane's major portion remained intact for a while after the separation of the red zone pieces and the nose. This portion of the airplane included both wings, most of the wing center section, and a small amount of fuselage structure in front of the wing center section. Aerodynamic calculations and radar data showed that the major portion of the airplane climbed and rolled and then began a steep descent to the water. As speeds and loads increased during the descent, the wing center section broke apart near the left wing. By examining the fire damage and soot patterns concentrated on the fuselage close to the right wing, the investigators concluded that an explosion must have occurred at this point because of fuel leaks from the right wing fuel tank. ([13], p.109)

In conclusion, the Sequencing Group found that the breakup of the TWA 800 aircraft began with the fracture of a forward boundary member of the center wing fuel tank—i.e., the spanwise beam 3—because of an overpressure event within the fuel tank. However, what caused the overpressure event was yet to be determined.

6.4 Ruling Out Alternatives

At the early stages of the investigation, three leading theories about the causes of the inflight breakup quickly emerged: missile attack, bombing, and massive structural failure and decompression of the 25-year-old plane. (The NTSB report grouped the first two theories under the category of "high energy detonation devices".) Flight 800 was among the oldest 747s in service, but the only previous 747 explosions involved bombing. ([6], p.60) When enough wreckage was recovered and examined, the investigators determined that these early theories could be ruled out.

6.4.1 Missile Strike

Because of rumors of a missile strike and possible eyewitness reports supporting these rumors, The FBI and the NTSB studied the missile theory extensively. The missile theory investigation focused on two major questions: Whether there is any physical evidence of a missile strike on the wreckage, and whether the radar data is consistent with a missile intersecting the trajectory of the airplane. The answers to these two questions turned out to be no.

First, blast patterns from a missile strike are well understood, and the investigators conducted additional testing on the TWA 800 wreckage to dispel any rumors. Testing showed that if a missile with a live warhead impacted the airplane and detonated, the wreckage would likely have exhibited extensive damage to the impact area from the initial penetration of the missile and the subsequent dispersion of high-velocity post-detonation fragments. For instance, a blast strong enough to penetrate the fuselage would produce petaling of the surface, pitting of adjacent surfaces from small blast fragments and hot gas surface "washing". If any of these effects were present on the wreckage, they should be clear to experienced investigators. Even after the recovery of 95% of the airplane wreckage, the investigators found no evidence of such blast effects. ([6], p.68)

Besides a direct missile strike, the investigators also considered the possibility of a missile fragment penetrating the fuel tank. The basic idea of this scenario is that instead of directly impacting the surface of the airplane, a shoulder launch missile might have self-destructed near the wing center section; some of its fragments penetrated the center wing fuel tank and caused an explosion. However, according to the missile tests, a shoulder launch missile would have to detonate within 40 feet for a fragment to penetrate the aluminum skin. At that distance, numerous other distinct high-velocity fragments would leave telltale impact marks in a starburst pattern. Again, no such blast pattern was found. ([6], p.68-69)

The investigators further examined 196 relatively small holes on the reconstructed fuselage to see if they have characteristics of high-velocity penetrations. Once again, the characteristics of high-velocity impact hole are well understood. Nevertheless, the investigators performed specific tests prepared by Boeing. The metallurgical characteristics of a high-velocity impact include: Splashback of materials on the perimeter of the hole on the entry side; melted and re-solidified materials on the wall next to the entry side; breakout deformation of material on the perimeter of the hole on the exit side. Two holes in the wreckage showed evidence of high-velocity impact, but they were made by objects traveling from inside to outside. ([6], p.64) In sum, no characteristic physical evidence of the detonation of a missile was ever found in the recovered wreckage pieces.

Second, to determine whether there was any evidence of a missile trajectory that intersected TWA flight 800, NTSB did a thorough review of recorded data at the time of the accident from long range and airport surveillance radars. Recorded radar data from ground-based radar antenna sites can be either primary or secondary. A primary radar target is recorded when a primary radar signal reflects off an object's surface and returns to the site for processing and display. A secondary radar target is recorded when a radar signal is detected by the airplane's transponder, which transmits a coded message back to the radar. Secondary radar returns contain airplane identification and altitude data for air traffic control purposes. The radar data reviewed by NTSB contain both primary and secondary radar returns. ([13], p.88)

Using the recorded radar data, NTSB investigators tracked TWA flight 800's flight path based on its secondary radar returns from the time the airplane departed JFK until its last secondary radar return. The radar data review also revealed multiple sets of primary and secondary radar returns from other airplanes or objects. None of the sequences of radar returns intersected TWA flight 800's position, nor are any radar returns consistent with a missile traveling towards the airplane. For instance, there are some unidentified primary radar tracks within 5 nautical miles (9.2 km; 5.7 mi) of TWA 800's flight path, but the speeds of these unidentified objects were too slow to be that of a missile. In short, no radar returns were consistent with a missile strike scenario. ([13], p.88-89)

Moreover, the NTSB addressed allegations based on the recorded data from the Long Island MacArthur Airport, Islip, New York. According to the allegations, data from this radar site showed groups of surface targets converging suspiciously near the accident. Further, a 30-knot radar track, never identified, and 3 nautical miles (5.6 km; 3.5 mi) from the crash site, was alleged to be in foul play, because it did not divert from its course to help with the search and rescue operations. However, military records showed no military surface vessels within 15 nautical miles (28 km; 17 mi) of the accident aircraft. Also, the records showed that the closest area scheduled for military use was 160 nautical miles (296 km; 184 mi) south. ([13], p.93)

The NTSB examined the 30-knot target track to determine why it did not divert from its course and return to where the TWA 800 wreckage had fallen. The investigators came up with the following explanation: TWA 800 was behind the target. Since the perspective of the target's occupant(s) was likely forward-looking, the occupants would not have been able to observe the aircraft's breakup or subsequent explosions or fireball(s). Moreover, it was unlikely that the occupants of the target track would have heard the explosions over the engine sounds and the boat's noise traveling through water, especially if the occupants were in a cabin. Finally, a review of the Islip radar data for other similar summer days and nights in 1999 showed that the 30-knot track was consistent with normal commercial, recreational, or cargo vessel traffic. ([13], p.94) In conclusion, the alleged anomalous radar returns recorded by the Islip, New York, radar site can be explained away without appeal to the missile strike scenario.

6.4.2 Bombing

Bombing was another prominent theory at the initial stages of the investigation and was taken seriously by the FBI investigators. A terrorist bomb can cause an in-flight breakup. However, bomb detonation damages have very distinctive patterns, and they were not present on Flight 800. For instance, the initial shock blast of high-velocity gas will shatter and disintegrate material immediately opposite the explosive charge, creating a blast hole or shatter zone. Other distinctive characteristics of detonation patterns include starburst fracture pattern, petaling, and, if the detonation occurs in the passenger cabin, ripped suitcases and shards of metal embedded in passengers. The investigators found none of these blast characteristics on recovered TWA Flight 800 wreckage, including the most obvious places to hide a bomb such as the cockpit and the passenger cabin. Only about 5 percent of the airplane's fuselage was not recovered. None of the missing fuselage areas were large enough to have included all the damages caused by a bomb's detonation. ([6], p.66-67)

The only evidence suggestive of a bomb was explosive residue on Flight 800. Just seven days after the crash, the FBI found a minimal amount of explosive residue on three separate pieces of aircraft wreckage. ([13], p.258) Later, when there were no other physical signs consistent with a bomb exploding, the investigators began considering the question: Where had those explosive residues come? A search of FAA (Federal Aviation Administration) records showed that in the spring about a month before the accident, at St. Louis' Lambert Airport, a dog-training explosive detection exercise included placing and removing explosives from several locations in the accident airplane. ([13], p.259) This discovery seemed to resolve the question about the source of the explosive residues satisfactorily.

However, further testing by the FAA found that the explosive residues on the TWA 800 would dissipate completely after two days of seawater immersion. The investigators recovered few pieces of the wreckage in the first two days after the accident, and those recovered pieces were floating in the ocean. The pieces with the explosive residues were likely immersed in ocean water for over two days before their recovery. ([13], p.259) Therefore, the dog-training exercise was unlikely to be the source of the explosive residues.

The investigators could not determine the exact source of explosive residue found on the wreckage. However, the lack of evidence corroborating a high-energy explosion showed that these residues were not the result of a high-energy explosive device's detonation on TWA flight 800. The most likely sources for the explosive residues, according to NTSB, were the ships or ground vehicles used in the recovery operations. Military personnel, ships, and ground vehicles used in the recovery operations had come into frequent contact with explosives. None of them were cleaned up to prevent contamination of the wreckage. Therefore, explosive residues could have been transferred from the surfaces of the ships or vehicles onto the wreckage during the recovery operations. ([13], p.259)

6.4.3 Structural Defects

Finally, the investigators also considered the possibility of a structural failure resulting in massive decompression and disintegration of the airplane. However, a close examination of the wreckage revealed no evidence of preexisting airplane structural failures (e.g., corrosion, fatigue, or mechanical damage) that could have contributed to the in-light breakup. For instance, the preexisting corrosion damage to the aircraft structure was minimal and could not have led to or affected the airplane's breakup. Similarly, small fatigue cracks were found in some parts of the airplane, but none of these cracks had grown into a propagating crack that could have led to the in-flight breakup. Therefore, the NTSB concluded that it was not a structural failure that initiated the in-flight breakup of TWA flight 800. ([13], p.257)

6.5 The Fuel/Air Explosion in the Center Wing Tank

The Sequencing Group determined that the TWA 800 breakup sequence was initiated by an overpressure inside the center wing tank (CWT). There was no evidence that a high-energy explosive device detonated in any area of the airplane, nor was there any evidence that a structural failure and decompression initiated the breakup. Therefore, NTSB concluded that the overpressure could only have been caused by a fuel/air explosion in the CWT. ([13], p.261)

However, because fuel-tank explosions are extremely rare occurrences, NTSB's conclusion generated puzzlement and concern for both investigators and the aircraft industry. To further examine the fuel/air explosion scenario, the investigators focused on three major questions. First, was the fuel tank explosive at the time of the accident? Second, could a fuel/air explosion generate enough pressure to break apart the fuel tank and destroy the airplane? Third, what was the ignition source? In the end, the NTSB obtained positive answers to the first two questions but could not arrive at a definite answer to the third. ([6], p.270)

6.5.1 Fuel Tank flammability and Combustion Research

During its investigation of the fuel/air explosion scenario, the NTSB recognized the lack of available relevant research data concerning the flammability of Jet A fuel[1] and its behavior in airplane fuel tanks. Because of that, NTSB carried out a research program to develop relevant data. The objectives of this research program included: (1) To develop an understanding of the thermal and vapor environment within the 747 center wing tank (CWT); (2) To determine the chemical and physical properties, flammability, and combustion behavior of Jet A fuel; (3) to develop computer models of the combustion process within the 747 CWT to help with the determination of the ignition location. ([13], p.138)

---

[1] The Jet A fuel is a fuel commonly used in jet transport fleet at the time.

For a fuel/air explosion to happen in a tank, the ratio of fuel molecules to air molecules in the vapor space[2] of the tank must be within a specific range. When a fuel partly fills the tank, it evaporates into the vapor space until it reaches an equilibrium. If the fuel-to-air ratio is too low, combustion is not sustainable because the distances among the fuel molecules are too far. Conversely, if the fuel-to-air ratio is too high, combustion cannot happen because there is not enough oxygen. The range of fuel-to-air ratios required for ignition and rapid combustion depends on various factors, including the fuel type, the temperature within and near the fuel tank, and altitude. ([6], p.271)

The type of fuel that the TWA flight 800 used is known as Jet A fuel. Jet A fuel is a derivative of kerosene that became widely used in the transport fleet because it was less flammable than earlier fuels. Fuels used in jet transport fleet must satisfy two needs: First, the fuel needs to burn well in jet engines and produce the thrust efficiently under a variety of conditions. Second, the fuel should not be too volatile; it should not evaporate too easily under normal or higher temperatures. Jet A fuel provided the best compromise between the two needs than earlier fuels used in the industry. ([14])

Based on flight record, the investigators knew that TWA Flight 800 left the JFK airport with a nearly empty center wing fuel tank, which contained approximately 300 pounds of residual Jet A fuel. The fuel would form a small pool, and the remaining space of the center wing tank would be the vapor space. ([6], p.271) Initially, the investigators knew very little about the thermal and vapor environment within the tank during the accident; thus, they did not have credible means to assess the CWT's flammability. ([14])

Therefore, in the summer of 1997, the Safety Board developed an experimental flight test program to get data concerning the thermal and vapor environment within and around the center wing tank. The experimental test program comprises a series of flight tests intended to emulate the accident flight conditions as closely as possible; for instance, the tests were conducted in July and very similar weather conditions to that of the accident flight. The

---

[2] The vapor space is the space in the fuel tank occupied by fuel vapor and air, and not by the liquid fuel.

investigators installed measurement sensors in the flight test airplane within and around its center wing tank. They continuously took temperature and pressure measurements during the entire period of ground and flight operations at nearly 200 different locations. Finally, they took fuel vapor samples from the tank. This flight test program was the first time that the thermal and vapor environment inside an in-flight aircraft fuel tank was measured. ([14])

The NTSB used multiple independent measures of flammability to assess whether the center wing tank on board the accident flight was flammable. First, test flights measured the temperature of the vapor space when test airplanes reached 13,800 feet, the altitude at which the accident happened. The results showed that the temperature of the vapor space had increased from its initial preflight temperatures (average less than 80 degrees Fahrenheit) to an average temperature of about 120 degrees Fahrenheit. All the fuel tank temperatures measured at 13,800 feet are higher than the "flash point" temperature of Jet-A fuel, which is the lowest temperature at which the fuel gives off vapor in sufficient concentration to combust when exposed to a standard ignition source. These temperature measurements showed that a flammable condition existed in the center wing tank of TWA 800 at the time of the accident. ([13], p.125-128)

Second, the NTSB also sought another measure of flammability by sampling the fuel vapors in the center wing tank on the test airplane, and then directly analyzing the chemical composition and concentration of the fuel vapors. The analysis of these samples also showed that the fuel vapor inside the center wing tank of the test airplane was flammable at 13,800 feet. ([13], p.129-130)

In sum, multiple analyses independently confirmed that the fuel-air vapor inside the center wing tank of Flight 800 was flammable. Given the relatively high flash point of the Jet A fuel, the fuel vapor's flammability was primarily because of the high temperature within the fuel tank. Based on this information, investigators examined the 747 center wing tank and its relationship to nearby components to determine how and why the temperatures in the center wing tank can rise to such high levels during the operation.

Beneath the center wing tank is the so-called pack bay, an enclosed space containing the air conditioning units called packs. The NTSB investigators realized that when the air conditioning packs were in use, waste heat from these packs raised the air temperature in the pack bay beneath the center wing tank. Flight tests emulating that of Flight 800 showed that the surface temperatures of some pack components might have exceeded 300 degrees Fahrenheit, and the air space in portions of the pack bay might have exceeded 200 degrees Fahrenheit before takeoff. These elevated temperatures drive heat into the center wing tank and raise the fuel vapor's temperature within the center wing tank. ([13], p.126)

A portion of heat in the center wing tank could escape to the air outside the airplane through the neighboring wing tanks and then through the wings. Once the aircraft begins its flight, the colder air at high altitudes helps remove heat from the center wing tank. However, this heat removal process can slow down if the airplane remains on the ground and in hot weather for an extended period. Flight 800 remained at the gate for nearly three hours with the air conditioning system operating, which likely contributed to the fuel-air vapor's flammability.

The investigators established that a flammable condition existed on board Flight 800 at the time of the accident, and they identified a major source of heat to the tank. They then proceeded to determine the energy needed for igniting the fuel/air vapor, the combustion pressures that could develop from the ignition, and the rate at which the pressures could develop. Initially, technical information required to answer these questions again did not exist in the literature, so the NTSB contracted Cal Tech to conduct an extensive research program of Jet A fuel flammability and combustion behavior. ([13], p.131)

One area of Cal Tech's research was to determine the energy required to ignite Jet A fuel vapor. To achieve this, the researchers constructed specialized combustion test chambers and devised electronic circuits to precisely measure the energy needed to ignite the fuel/air vapors by a small spark. They ran several hundred experiments to determine Jet A fuel's ignition limits. These test results showed that for the conditions of Flight 800, the spark ignition energy required was between 0.5 and 500 millijoules (mJ). ([6], p.271) For reference, the energy in a static electric spark from walking on a rug is between 1 and 10 millijoules. ([6], p.278)

Another area of Cal Tech's research was to determine the combustion behavior (for example, explosion peak pressures and flame speeds) of Jet A fuel. For this aim, the investigators carried out two major types of experiments. They conducted the first major type of experiment in large combustion chambers under a wide variety of conditions. They measured the maximum overpressures that resulted from the combustion of Jet A fuel vapor and the speed at which this occurred. The experimental results showed that for conditions on TWA flight 800, the range of peak pressure is greater than the pressures needed to break apart Flight 800's center wing tank. ([13], p.131-133)

The second major type of experiments carried out by the Cal Tech researchers used quarter-scale models of the center wing tank. These quarter scale models are one quarter the length, width, and height of the actual full-scale center wing tank, and incorporate important features of the actual tank such as vents and passageways. The purpose of using these quarter-scale models is to investigate the combustion behavior in chambers that are more representative of the actual center wing tank, and to collect data towards the development of computer models of center wing tank combustion. The quarter-scale model testing results confirmed that the peak explosion pressures measured exceeded the calculated structural limits of the center wing fuel tank. ([13], p.134-136)

Finally, based on the data collected from flight tests, combustion chamber experiments, and quarter-scale model tests, the NTSB developed a computer modeling simulation program to determine the fuel tank ignition location for Flight 800. The strategy of the research program was to (1) use computer models to simulate the combustion processes within the center wing tank under a variety of scenarios (e.g., different ignition locations, different temperatures, variations in quenching (flames going out), etc.); (2) predict the structural damages in each computer simulation scenario; (3) compare the predicted damages to the observed damages in the wreckage of Flight 800. The hope was that if they could find the primary ignition location, they could identify the ignition source. ([13], p.137)

However, the center wing tank's geometry is very complex, and simulating the combustion in it was an arduous process. The investigators had to simulate the chemical combustion and the fluid dynamics of fuel vapors being pressurized and passing from one to another through vents and passageways. Combining all the factors was a challenging task, and it led to a high level of uncertainty throughout the calculations. In the end, the uncertainties in the computer modeling were too great to permit the identification of the ignition's probable location. However, the analysis found several ignition locations that predicted damages consistent with the damage observed in the wreckage and the structural failure calculations. ([13], p.137-139)

In summary, NTSB's fuel tank flammability and combustion research established two important conclusions: First, the Jet A fuel/air vapor inside the center wing tank of TWA Flight 800 was flammable. Second, the ignition and combustion of this flammable mixture of fuel/air vapor could generate sufficient pressure to break the center wing tank structure.

6.5.2 The Search for the Ignition Source

To find out what ignited the flammable fuel/air vapor in the center wing tank and caused the explosion, NTSB investigators considered many potential ignition sources and went through a process of elimination. Eventually, the NTSB concluded that the vast majority of the potential ignition sources were highly unlikely. The most likely ignition source was a short circuit event outside of the fuel tank that provided excessive voltage through the low-voltage fuel tank wiring. Other potential ignition sources that were unlikely included: a meteor strike, missile strike or bombs, auto-ignition or hot surface ignition; uncontained engine failure, turbine burst in air conditioning pack, malfunctioning fuel pumps, static electricity, and electromagnetic energy radiated from transmitters outside the airplane coupled to the fuel gauge wiring. ([6], p.282)

The investigators ruled out some potential ignition sources because their characteristic traces were not present in the accident airplane's wreckage. For instance, they ruled out missile strike and bomb scenarios because of the lack of characteristic damage patterns of high energy detonation. ([13], p.272-273) Another potential scenario ruled out in this way is auto-ignition[3] or hot surface ignition[4]. Auto-ignition of the fuel could occur at 460°F, and hot surface ignition could occur if any part of the metal fuel tank walls reached 900 to 1300°F. However, other types of thermal damage would be apparent if these temperatures were reached. Investigators considered various possible failures (such as a fire in the air conditioning pack bay beneath the center wing tank, or a fire in the main landing gear wheel) that might cause hot surfaces. They found no physical evidence consistent with these failures. ([13], p.274)

The investigators ruled out other potential ignition sources because they could not generate sufficient ignition energy to ignite the center wing tank, given the condition at the time of the accident. For instance, investigators theorized that electrically isolated parts in the CWT could become highly charged with static electricity (generated by fuel sloshing). A discharge from such a part could create a spark to ignite the fuel/air vapor in the CWT. However, in a laboratory setting designed to imitate the center wing tank of TWA 800, the highest voltage potential was 650 volts. The discharge energy such a voltage could produce was estimated to be at most 0.03 mJ, which is still well below the 0.25 mJ minimal ignition energy for Jet A fuel vapor. ([13], p.278) Similarly, the investigators ruled out radiated electromagnetic interference, because tests showed that it could not have generated enough energy to ignite the fuel tank. ([13], p.280)

Of all the potential ignition sources evaluated, the most likely was a short circuit outside the center wing tank that transferred excessive voltage to the tank through the fuel quantity

---

[3] Auto-ignition is spontaneous ignition without a spark or other ignition source.

[4] Hot surface ignition is ignition by being in contact with a localized hot surface.

indication system (FQIS, or fuel gauge system) wiring. The fuel gauge system includes probes located inside each fuel tank to measure fuel quantity. The only electrical wiring located inside the CWT is the wires connected to these fuel gauge probes. However, the fuel gauge wiring carried very low voltage (about 25 volts) and could not discharge enough energy to ignite the fuel tank. Therefore, if fuel gauge wiring had played a role in the CWT's ignition, the following two events must have occurred. First, some power source outside of the fuel tank transferred higher-than-intended voltage into the fuel gauge wiring. Second, the excess energy was released inside the tank and ignited the fuel/air vapor in the tank. ([13], p.279)

The fuel gauge wires connected the probes in the fuel tank and the fuel gauge in the cockpit. Even though the fuel gauge uses only a low level of power to operate, fuel gauge wires are routed in bundles with high-voltage wires that power other aircraft systems. The investigators theorized that a short circuit involving high-voltage wires and fuel gauge wires inadvertently transferred the higher voltage from the former into the latter. Although there was no smoking gun evidence conclusively establishing that such a short circuit occurred and initiated the fuel tank explosion, the investigators obtained a great deal of supporting evidence. Besides ruling out all other possible sources of ignition in the fuel tank, the following five points support the theory. ([6], p.283)

First, there is extensive evidence of wire degradation and damage, both in the accident airplane's wreckage and in 26 other transport category airplanes from various operators. Much of the wire recovered from the TWA 800 wreckage had damaged insulation and cracks that exposed the copper core conductors. A short circuit can occur if the wires' internal conductors are exposed and if there is direct contact between bare copper. Inspection of other in-service airplanes showed that (1) there is no uniform standard in the commercial aviation industry for proper wire separation between different wires in a wire bundle, and (2) damaged wire insulation is common among other inspected airplanes. A short circuit scenario is consistent with all these findings. ([13], p.282-283)

Second, a short circuit can also occur if there is a bridge between cracked wires by contaminants such as metal shavings or fluid, and the investigators found evidence of such contaminants in the accident airplane (and in other airplanes examined). For instance, airplane repairs and structural modifications often require extensive drilling that would have created drill shavings. The NTSB's Maintenance Records Group found that, rather than removing drill shavings, a common practice adopted by maintenance people was to use compressed air to blow shavings off the repaired structure. Also, the Systems Group found drill shavings on a cabin floor beam fragment two inches from the fuel gauge wires. The discovery of the drill shavings raised the possibility that similar shavings had gotten into the wire bundles, damaged the wires, and caused the short circuit. The number of repairs found throughout the wreckage showed many possible locations for a short circuit on TWA 800. ([14])

Third, the NTSB carried out a series of short-circuit tests involving damaged wires. The investigators studied three types of short circuits: (1) wet short circuits, in which lavatory fluids contaminated the wires with damaged insulation. (2) Dry, nonabrasive short circuits, in which metal shavings were placed between wires with damaged insulation. (3) Dry, abrasive short circuits, in which metal shavings were placed between vibrating wires with intact insulation. These tests showed that contaminants bridging damaged wires could create short circuits that lasted for many minutes. The tests showed that peak currents of over 100 amps of a parent wire could release up to 400 mJ of energy to a lower voltage victim wire. A short circuit occurred, even if circuit breakers protected the victim wire. The 400 mJ of transferred energy greatly exceeded the 0.002 mJ energy normally supplied to the fuel gauge wiring and the smallest experimentally measured ignition energy of 0.5 mJ. ([6], p.284)

The above three points support the claim that excess energy could have entered the center wing fuel gauge wiring through a short circuit. The short circuit could have been created by direct contact between exposed wire conductors, or by indirect contact with metal shavings or fluid. The next point concerns how this excess energy might have been released in the tank.

Fourth, even though the investigators only recovered a small amount of wiring from the accident plane's center wing tank, they found multiple possible energy release mechanisms via the fuel gauge wiring. Wires within the fuel tank are made of silver-plated copper, and pre-accident damage to the exposed copper core of wires existed in the wing tanks of TWA 800 and the fuel tanks of other airplanes. One possible energy release mechanism within the CWT is that voltage from the exposed conductors arced to other metal, such as other nearby damaged wires, or the metal parts of the center wing tank. ([13], p.290)

Another possible energy release mechanism within the CWT is through the sulfide deposits found on the accident airplane's fuel gauge parts. The NTSB found a thin film of sulfide deposits on a fuel probe connector in the airplane's wreckage. It resulted from an interaction of sulfur contaminants in the fuel with the silver-plated fuel gauge wires within the CWT. Similar deposits also existed on the fuel gauge probes and probe wiring from the fuel tanks of other airplanes. Sulfide deposits are semi-conductive. When the investigators applied a 170 volts-pulse to the wiring with sulfide deposits, the deposits burned off in a bright flash and a loud pop. Further laboratory tests showed that the application of direct current voltage to sulfide deposits could ignite Jet A fuel vapor. ([13], p.291)

So far, the list of evidence given above only shows that a short circuit *could have* transferred excess energy through the fuel gauge wires and that the energy *could have* been released inside the tank via several possible mechanisms. It is more difficult to show that a short circuit *actually happened* and *actually caused* the ignition of the fuel/air vapor, because the accident obliterated much of the physical evidence needed. For instance, even though there is evidence of arcing and melted conductors in some fuel gauge wires recovered from the accident airplane, the arced wires all contained fire damage. The investigators could not determine if arcing took place before or after the fuel tank explosion. Moreover, the NTSB was only able to identify about half of the original amount of fuel gauge wiring.

So, there easily might have been other locations where a short circuit could have taken place. Nevertheless, the next point shows that there is some indirect evidence for a pre-accident short-circuit.

Fifth, there is evidence of multiple electrical anomalies on Flight 800 right before the accident happened. About 2 and 1/2 minutes before the cockpit voice recorder on TWA 800 lost power, the captain commented about a "crazy" fuel flow indicator, which suggests that some electrical anomaly occurred that affected the fuel flow indication system wiring. Fuel flow indication system wiring is routed in a common wire bundle with fuel gauge wiring and high voltage wiring that power cabin lighting. ([13], p.289) Moreover, maintenance records showed that a pre-accident structural modification was made at a location very close to the wire bundle that includes both fuel gauge wiring and fuel flow indication wiring. The structural modification involved an extensive amount of drilling. ([14])

In addition, the cockpit voice recorder registered a few interruptions in the background noise about two seconds before power loss. The interruptions indicated a brief drain on the electrical power to the cockpit voice recorder, which implies that some electrical anomaly in the adjacent circuits was drawing a high amount of current at the time of the interruptions. ([13], p.289)

Finally, the recovered CWT fuel gauge from the cockpit displayed a reading of 640 pounds of residual fuel, which does not agree with the quantity recorded by the ground refueler (300 pounds). The NTSB testing showed that applying a high voltage to fuel gauge wires can cause the cockpit fuel gauge digital display to change by several hundred pounds in less time than needed to trip the circuit breaker. It follows that an electrical anomaly involving fuel gauge wires might have affected the reading of the cockpit fuel gauge. ([13], p.290)

The NTSB investigators acknowledged that these electrical anomalies might not have the same cause. However, at least one of these anomalies was likely the result of an electrical event that also transferred an excess voltage to the CWT FQIS (fuel gauge) wiring. Based on these considerations, the Safety Board concluded that the most likely ignition source for the center wing tank explosion was a short circuit that transferred excessive voltage to the CWT through fuel gauge wires. ([13], p.290)

6.6 Analyzing Witness Reports

As mentioned earlier, many witnesses near the accident reported seeing a streak of light followed by a fireball in the sky. After the accident, there was much speculation among the public that the witnesses saw a missile that struck TWA flight 800. However, because the physical evidence ruled out the possibility of a missile strike, the witnesses must have observed something other than a missile. To determine what these witnesses were observing, NTSB investigators formed a Witness Group to study all the witness interview summaries documented by the FBI. ([13], p.262)

The Witness Group found that the FBI documents contained 736 witness accounts. Out of the 736 witnesses, 258 were "streak-of-light" witnesses, meaning that they saw an object in the sky moving like a streak of light, a flare, or something similar. Of 258 streak-of-light witnesses, 38 reported that the streak was ascending almost vertically, and another 18 reported that it originated near the horizon. Moreover, 599 out of the total 736 witnesses were "fireball" witnesses. Out of the 599 fireball witnesses, 264 reported seeing the fireball originate, 200 reported seeing the fireball split into two fireballs, and 217 reported seeing the fireball hit the water (or disappear below the horizon). Finally, 210 witnesses reported seeing both a fireball and a streak of light. ([13], p.230-232)

To determine what the streak of light and the fireballs were, the Witness Group relied on the breakup sequence of the accident airplane, which was reconstructed based on sequencing study, computer simulation and timing information provided by witnesses:

When the accident airplane was in flight near the Long Island Coast, an explosion of the center wing tank had occurred. About 3-5 seconds later, the nose section departed and fell to the water. However, the rest of the airplane—including most of the wing center section, the wings, the aft fuselage, and the tail—continued to fly. Computer simulations showed that the crippled airplane ascended from 13,800 to about 15,000-16,000 feet, and then went into a descending turn to the right. Shortly after the descending turn (over 30 seconds after the center wing tank explosion), the outboard portions of the wings separated. Soon after, the wing center section separated near the left wing, causing the left wing to separate from what remained of the airplane. A water impact ensued. The entire breakup sequence took about 47 to 54 seconds. ([13], p.263-264.)

Based on this breakup sequence, the NTSB investigators considered what the whole breakup sequence might have looked like to someone watching from Long Island. First, the investigators identified the locations of nearly all the witnesses at the time of the accident. About three-quarters of the witnesses were 11 miles away or further, and nearly 100 witnesses were over 23 miles away. From those distances, the whole airplane would have been tiny and barely noticeable at all. The explosion of the center wing tank occurred inside the intact aircraft, and it was unlikely that witnesses would see this explosion. The nose section separated a few seconds after the CWT explosion, and the wreckage of the nose portion was virtually free of fire or heat damage. In other words, the nose section was not illuminated during its fall, so it was also unlikely that witnesses would have seen it. ([15])

After the nose portion's separation, a fuel-fed fire in the exposed CWT would likely have been visible to witnesses some distance away. Such a fire would have looked like a small streak of light. When the outboard portions of the wings separated, it opened up the outboard wing fuel tanks, which likely led to a growing fire. To a witness, this could have resembled an explosion followed by a fireball. As the aircraft continued to disintegrate, the fireball would grow in intensity. Shortly after the outboard portion of the wings separated, the remaining left wing and the wing center section separated from the rest of the airplane. The development of a fire associated with these failures would probably appear as a "splitting" of the fireball, which would have appeared to disappear from view below the horizon, trees, or other obstructions. ([13], p.263-264)

In sum, if a witness could see the entire breakup sequence, what we would expect them to see was a streak of light followed by a fireball, and the fireball might split into multiple fireballs. This conclusion is consistent with the descriptions of the vast majority of the witnesses. The NTSB further examined the accounts of a few prominent witnesses (e.g., the captain of Stinger Bee flight 507) in greater detail, because of their unique vantage points or the level of precision and detail in their accounts. Analysis of these selected accounts showed that they are consistent with the witnesses having observed some part of the breakup sequence after the CWT explosion. ([13], p.264)

Although most of the observations reported in the witness documents were consistent with the witnesses having observed some portion of the airplane's crippled flight following the explosion of the CWT, the NTSB could not explain a small percent of the reported witness observations in this way. In particular, 38 witnesses described a streak of light ascending vertically or nearly so. An additional 18 witnesses reported seeing a streak of light originating at the surface or from the horizon. These 56 accounts appeared to be inconsistent with the accident airplane's calculated flight path and other aspects of the accident sequence, if the streak of light were the crippled airplane. ([13], p.265)

To reconcile the calculated accident sequence with these 56 witness reports, the NTSB argued that the anomalous witness reports could be explained away as the effect of potential deficiencies in the interviewing and documentation process and errors in witness perception or memory.

First, the NTSB pointed out that the FBI conducted almost all witness interviews during its criminal investigation, aiming to determine whether a missile struck TWA flight 800. The criminal investigation's focus was apparent in the suggested interview questions the FBI provided its agents, which presuppose a missile attack. For instance, some suggested interview questions were: "How long did the missile fly?", "What does the terrain around the launch site look like?", and "Where was the sun in relation to the aircraft and the missile launch site?". ([13], p.266) The NTSB cited psychological research to argue that the framing of the interview questions and the witnesses' deference to the FBI agents' expertise could have biased the interviewee's answers towards the missile strike scenario. ([13], p.266)

Second, the NTSB also criticized the FBI's documentation of the witness accounts. Instead of keeping verbatim records of the witness interviews, the FBI agents wrote summaries of the interviews, and the witnesses did not review these summaries. The lack of an accurate and reliable documentation process could introduce errors concerning the origin and trajectory of the streak of light. For instance, at least three streak-of-light witnesses stated during re-interviews that, contrary to the FBI documentation of their earlier interviews, they did not observe the streak of light originating from the earth's surface. ([13], p.267)

Third, the NTSB suggested that the anomalous witness reports could result from various perceptual illusion. One possibility is that some witnesses might have been able to see the crippled airplane coming nearly directly toward them, which could have resulted in the appearance of a nearly vertical rising streak of light. Another possibility is the so-called equidistance tendency, which is the tendency to perceive more distant objects as being about the same distance away as intervening objects. Given the weather condition at the time of the accident, witnesses observed the accident from Long Island would have been viewing it against a featureless background and, therefore, would have had few depth perception cues. For some observers, the equidistance tendency would have caused the streak of light to appear lower in the sky and closer to some intervening terrain features, such as houses or the barrier island. It would explain why some witnesses may have reported the streak of light as having originated from the surface. ([13], 267-268.)

Fourth, the NTSB cited psychological research to argue that witnesses might have unintentionally created inaccurate memories because of exposure to post-accident information. Some witnesses' recollections might be distorted because of exposure to other witness accounts and media reports. If many people believed that a missile struck the airplane, the witness might eventually report the misinformation as a memory. As an illustration, the documents recording the initial interviews of some witnesses did not report a missile; however, the same witnesses later stated that a missile strike happened. Certain misleading details such as a vertically rising streak of light might also have been assimilated as false memories. ([13], p.269)

In sum, because of all these possible ways of explaining away the 56 abnormal witness reports, the NTSB did not regard these abnormal witness reports as conclusive evidence that some witnesses had observed a missile. Therefore, NTSB concluded that the witnesses did not see a missile strike, but some part of the breakup sequence following a CWT explosion.

6.7 Conclusion

The NTSB concluded that the first event of the TWA 800 in-flight breakup was an explosion of the center wing fuel tank (CWT). The investigators did not find any physical evidence of a bomb or missile strike, but they found evidence of an overpressure event inside the center wing tank. Researches and tests showed that a Jet A fuel/air vapor explosion in the center wing tank caused the overpressure event. The investigators could not determine the source of ignition energy for the explosion with certainty. Of the sources examined by the investigators, the most likely was a short circuit outside of the CWT that transferred excessive voltage to the CWT through electrical wiring associated with the fuel gauge system. ([13], p.308)

The NTSB also identified many contributing factors to the accident, and it singled out two of them as prominent. The first significant contributing factor was the design and certification concept that fuel tank explosions could be prevented solely by excluding all ignition sources. The NTSB argued that such an approach was seriously flawed. The examination of many potential ignition sources had shown that there was no reliable way to eliminate all ignition sources. Nor was it rational to believe that we could predict all ignition sources. Instead, the most effective way to prevent fuel tank explosions should be eliminating flammable vapors inside the fuel tanks, combined with eliminating as many ignition sources as possible. ([14])

The second significant contributing factor was the design and certification of Boeing 747 with heat sources beneath the CWT. The results of the flight tests and the flammability research showed that many commercial aircraft might routinely have flammable fuel/air vapor in their fuel tanks. Airplanes with air conditioning packs located directly beneath their center wing tank are especially likely to have flammable vapor in the tanks. ([14])

Based on its conclusions, the NTSB issued fifteen safety recommendations, many of which were about fuel tank related issues. The NTSB's major conclusion was that the TWA flight 800 accident would not have occurred if there had not been flammable vapor in the center wing tank. Consequently, it recommended that the FAA and the aircraft industry give serious considerations to design modifications that would make fuel tanks nonflammable, such as inerting fuel tanks with nitrogen. ([13], p.309-312)

Part II

Reverse Causal Inference

Chapter 7

Introduction to Part II

The past events that researchers are interested in reconstructing are typically complex, in the sense that they are composed of numerous smaller, causally connected ‘subevents’. To reconstruct a past event of interest, it is necessary to identify the subevents that constitute it and establish the causal relationships among these subevents. Therefore, an account of the methodology of event reconstruction research cannot be complete without including an account of *causal inference* used in this type of research.

What is an inference? I define an inference as a step-by-step reasoning process that either discovers, or justifies, certain conclusions.[1] For instance, when I use Dijkstra’s shortest path algorithm to identify the shortest path between two vertices in a particular graph, I am carrying out an inference that helps me discover a particular conclusion. Moreover, when I produce a correctness proof of Dijkstra’s algorithm to argue that the path found by the algorithm is indeed the shortest, I am carrying out an inference that helps me justify that conclusion. Some inferences only play a discovery role, and some inferences only play a justificatory role. However, some inferences play both roles, and it is not always possible

---

[1] The phrase “discover a conclusion” is a bit awkward, it may be more natural to say either “draw a conclusion” or “discover the phenomenon expressed by a conclusion”. I use the term “discovery” mostly because it provides a familiar contrast with the term “justification” for philosophy audience.

to cleanly separate the two roles played by the same inference.[2]

A causal inference is an inference with one or more causal conclusions. A causal conclusion is a statement that expresses a causal relationship, which could be at the type level or the token level, and whose causal relata could be events, conditions, state of affairs, etc. There are many subtypes and distinctions within the category of causal inference. To understand the methodology of event reconstruction research, a particularly relevant distinction is the distinction between forward causal inference and reverse causal inference.

The main difference between forward and reverse causal inference is the *direction* of the inference. A forward causal inference infers from an event or condition (or a collection of events and conditions) to its effects on something else. A reverse causal inference, in contrast, infers from an event or condition to its causes. Another way to characterize the difference between the two types of causal inference is that they have different tasks. The task of a forward causal inference is to determine the effects of a given event or condition. In contrast, the task of a reverse causal inference is to determine the causes of a given event or condition.[3]

Although both forward causal inference and reverse causal inference are used in event reconstruction research, reverse causal inference plays a particularly prominent role, which is not surprising given that event reconstruction research is fundamentally backward-looking. After all, the event reconstruction researchers reconstruct what happened in the past based on the traces they have now. Therefore, I focus primarily on the methodology of reverse causal inference in this part of the dissertation, although I also discuss a hybrid method (process tracing) that combines characteristics of both forward and reverse causal inferences.

---

[2] The three types of causal inference that I will discuss—feature dependence, additional outcomes, and process tracing—all play both roles. Feature dependence and additional outcomes are *arguments*, so they are *primarily* justificatory, but they can also help discover certain causal conclusions. Process tracing, in contrast, is primarily a discovery procedure, although it can provide partial justifications of its conclusions.

[3] For a methodological discussion of forward versus reverse causal inference in statistics, see Andrew Gelman and Guido Imbens,“Why ask Why? Forward Causal Inference and Reverse Causal Questions” ([23]).

Finally, to clarify the connection between event reconstruction and reverse causal inference, it is helpful to distinguish between two types of epistemic situations that the investigators could be in when carrying out a reverse causal inference. Let event E be a given subevent within a larger event Σ of interest, and suppose the task of the reverse causal inference is to identify the causes of E.

The first epistemic situation is one in which the investigators know all the other subevents of Σ in some sense[4], and their task is to determine which of the other known subevents made causal contributions to the given subevent E. This epistemic situation rarely obtains in the practices of real event reconstruction research, and it occurs mostly in artificial settings. For instance, when philosophers apply a philosophical theory of causation to a self-contained imaginary scenario, the relevant subevents within the scenario have been fully postulated and described.

The second epistemic situation is one in which the investigators do not know many or most of the subevents of Σ. For the subevents they do know something about, they may not know *enough* about those events: For instance, they may know that an event of a certain type occurred, but not know any features of that event beyond its type. This type of epistemic situation occurs very frequently in event reconstruction research; consequently, the reverse causal inferences used in these practices often accomplish two tasks: To determine what subevents had occurred, and to determine the causal relationships among these subevents. Moreover, the task of reconstructing the past events and that of identifying causal relationships are often intermixed, as we shall see when we examine reverse causal inference methods such as feature dependence and additional outcomes.

This part of the dissertation is organized as follows. In Chapter 8 ("Events, Features, and Traces"), I clarify three major concepts that are used throughout the rest of this dissertation—events, features, and traces. I begin by making explicit a few metaphysical

---

[4] For instance, the investigators may know about the existence, the types and at least some of the features of these subevents.

assumptions about events. Then I introduce the concept of "features of an event" and illustrate three main types of features with examples. Finally, I propose an account of what traces are and how they provide epistemic access to features of past events.

In Chapter 9 ("Feature Dependence"), I introduce the first type of reverse causal inference that I call "feature dependence" or "feature dependence arguments". This type of inference makes uses of *feature dependence statements*, which state that features of an event depend on features of its causes. After discussing the basic forms and subtypes of feature dependence statements, I propose an account of the evidential basis of feature dependence statements, using the idea of *event-based causal systems.* Next, I articulate the structure of a type of feature dependence argument called "descriptive enrichment", the basic idea of which is that we combine multiple features of an outcome to construct a detailed description of one of its causes. Finally, I show how descriptive enrichment is used in practice, using an example from the American Airlines Flight 191 investigation.

In Chapter 10 ("Additional Outcomes"), I introduce the second type of reverse causal inference that I call "additional outcomes" or "additional outcomes arguments". The basic idea of this type of inference is that we can evaluate whether a hypothetical event C caused a given primary outcome E, by examining whether C caused an additional outcome E^*. I distinguish between two versions of additional outcomes and articulate their structures: A negative version, which concludes that C was not a cause of E; and a positive version, which concludes that C was a cause of E. Because the structure of the positive version is complex, I end the chapter by illustrating it with two full examples.

In Chapter 11 ("Process Tracing"), I discuss a hybrid causal inference method that combines elements of both forward and reverse causal inference. Process tracing establishes a causal connection between two events by tracing sequences of causally connected mediating events. Even though it is not a reverse causal inference *per se*, it complements other reverse causal inferences such as feature dependence and additional outcomes, and also serves as a natural transition to the next part of the dissertation on coherence and narratives. In this chapter, I propose an account of process tracing using Mackie's concept of INUS conditions, and a major part of the chapter consists of a *very* large example of process tracing, again from the American Airlines Flight 191 investigation.

Finally, I conclude this part of the dissertation with a recap of the major findings, and a few comments about why (and the sense in which) reverse causal inference tend to be more "successful" in engineering investigations than in other types of event reconstruction research in history or historical sciences.

Chapter 8

Events, Features, and Traces

8.1 Events and Features

Given my focus on the methodology of event reconstruction, I would like to begin by clarifying what I mean by "event". Even though I am not committed to a specific metaphysical theory of events, I do make four major assumptions about events throughout this dissertation:

1. First, events have spatial and temporal boundaries.
2. Second, events (typically[1]) can be decomposed into smaller "subevents".
3. Third, for each event, certain entities (objects) are "involved in" it.
4. Fourth, each event has many different *features*, which allow the same event to be described in many different ways.

To understand feature dependence, the fourth assumption is the most important, and I will dedicate most of this section to explaining it. Before I do that, I will use the crash

---

[1] I say "typically" in order to avoid addressing the metaphysical question about whether there exist atomic events that could not be decomposed further. In practice, engineering failures are typically complex events and can be decomposed into smaller subevents.

landing of United flight 232 to illustrate the first three assumptions.

First, as an event, the crash landing of United flight 232 occupies a particular spatial and temporal region. The event's temporal boundary began at the first ground contact of the aircraft by the right wing, and ended when the final pieces of the aircraft (e.g., the fuselage center section) came to rest. The spatial boundary of the event is more difficult to define precisely, but we know that the event is located within Sioux Gateway Airport in Sioux City, Iowa. More specifically, the event occurred more or less along the runway 22, where the aircraft crash-landed. ([11], p.5) The precise shape of the spatial boundary of the event is defined by how different parts of the aircraft broke apart and the parts' trajectories until they all came to rest.

Second, the crash landing event as a whole can be decomposed into smaller subevents. Some of the subevents of the crash landing event include: (1) The ground contact of the right wing tip; (2) the ground contact of the right main landing gear; (3) the ground contact of the No.3 (right) engine; (4) the No.3 engine spilling fuel and catching fire; (5) the main fuselage catching fire; (6) the cockpit, the tail and the right wing breaking off from the main fuselage; (7) the fuselage breaking into three sections; (8) the center section rolling into an inverted position and coming to rest in a cornfield near the runway, etc. ([6], p.172) The spatial-temporal boundary of each subevent is inside the spatial-temporal boundary of the entire event, and different subevents may bear various spatial and temporal relationships to each other. For instance, two subevents may be concurrent (having more or less the same temporal boundary), temporally overlapping or temporally separated; they may occur at the same location or different locations, and so on.

Third, many entities (objects or things, including human beings) were "involved in" the crash landing event. I do not have a precise account of the term "involved in", but it is easy to illustrate with examples. Some entities involved in the crash landing event already existed before the event. For instance, the right wingtip, the right main landing gear, the right engine, and all the passengers on board the flight were involved in the crash landing, and they all existed before this event. Some entities involved in the crash landing event ceased to exist after the event. For instance, the fuselage no longer existed after the crash, nor did the 111 people on board who died from the crash. Finally, some entities involved in the crash landing event were created during the event. For instance, the numerous fracture surfaces formed during the crash landing were new entities created during the event.

The fourth assumption that events have features requires a more extended discussion. Intuitively, features of an event are details of how the event occurred. We can think of an event as a concrete being, and each feature of the event as a particular aspect of this concrete being. Beyond that, I do not have a metaphysical account of what features are.[2] Fortunately, to understand how feature dependence works as a type of reverse causal inference, a metaphysical account of features is unnecessary. Instead, it suffices to describe the main types of features commonly used in event reconstructions, which I will do in the rest of this section.

There are three main types of features commonly used in engineering failure investigations and other types of event reconstructions: First, *some* properties of the entities involved in an event are features of that event. Second, the identities of the entities involved in an event are features of that event. Third, the spatial and temporal characteristics of an event are features of that event. I do not claim that these three types of features exhaust all the possibilities of features, nor that the different types must be mutually exclusive.[3] The only claim I am making is that these three types of features are the most common features you will see when reasoning about the causes of engineering failures.

---

[2] My notion of "feature" may remind metaphysicians of L.A. Paul's notion of "aspect" in "Aspect Causation" ([29]). Although, as we shall see, the notion of feature is broader than Paul's notion of aspect, because features are not limited to the instantiation of properties. Moreover, arguably my notion of features encompasses metaphysically heterogeneous categories. If this is true, it is not a problem for my account, because my notion of features is not motivated by metaphysical concerns, but rather by a methodological need to capture patterns of inference in an area of inquiry.

[3] For instance, some philosophers may argue that the identity of an entity can be reduced to some properties of that entity, and so the second type of features can be reduced to the first. I remain non-committed as to whether any type of features can be reduced to another type of features, or whether different types of features overlap with each other.

Let us begin with the first type of features of an event. Suppose that the temporal boundary of an event E is $[t_1, t_2]$, and that an entity X was involved in the event E. Then I claim that *at least some* of the properties of X during the time span $[t_1, t_2]$ are features of the event E. Similarly, if multiple entities $X_1, \ldots, X_n$ were involved in the event E, then *at least some* of the relations among $X_1, \ldots, X_n$ during $[t_1, t_2]$ are features of E.

I want to make three general comments about this type of feature before illustrating it with examples. First, in order for a property P of an entity X involved in an event E to be a feature of E, X must have property P at some point during the time span $[t_1, t_2]$ of E. For instance, if X only acquired property P after the event E was over, then P is not a feature of E.

Second, I used the qualifier *at least some* in the description of this type of feature. It is possible to go further and argue that if an entity X was involved in the event E, then *all* the properties of X during the time span $[t_1, t_2]$ of E are features of E. However, such a position has some counter-intuitive consequences. For instance, it would entail that since the fuselage of United 232 was involved in the crash landing event, the color painted on the fuselage exterior was a feature of that crash landing event. This may be counter-intuitive because the color of the fuselage exterior seems to have nothing to do with the crash landing event. For this reason, I only make the weaker claim that *at least some* of the properties of X during the period $[t_1, t_2]$ are features of event E. I do not, however, have a general account of *which* properties of X during $[t_1, t_2]$ are features of E. For the examples I will discuss in this chapter, I rely on an implicit understanding of which properties of the involved entities are features of the event in question.

Finally, the claim that some properties of the entities involved in an event are features of that event may sound strange, since the primary bearers of those properties were the entities (objects) rather than the events. I agree that these properties are features of the event only in a secondary sense. However, it is also true that the (relevant) properties of the involved entities are details about how the event occurred, and they can be aptly described as features of the event in question.

To illustrate this type of feature of an event, I discuss two examples. The first example concerns what I call "fracture events", and I argue that the properties of the fracture surfaces formed during a fracture event are features of that event. The second example concerns an aircraft encounter with a mountain rotor, and I argue that the properties of the aircraft, the properties of the rotor, and the relationships between the two during the counter are features of that event.

First, a fracture event is an event in which a fracture (crack) is formed.[4] As understood in failure analysis, fracture events typically consist of three phases: crack initiation, crack propagation, and final separation of the parts. The fracture surfaces were being created during all the three phases, and were finally completed during the last phase. Since the fracture surfaces were new entities created during the fracture events, they were involved in the fracture events.

Some current[5] properties of the fracture surfaces were created after the fracture event was over. For instance, after the TWA 800 disintegrated mid-air and crashed into the Atlantic Ocean near East Moriches, New York, the newly created fracture surfaces were immersed in the seawater for days, and the fracture surfaces suffered from some corrosion damage. The corrosion damage is a current property of the fracture surfaces, but they are not features of the fracture events because they were created outside of the event's time span.

For properties of the fracture surfaces created during the fracture events, at least some of them (if not all) are features of the fracture events. Some properties of fracture surfaces are macroscopic and can be observed using the naked eye: For instance, the size and shape of the fracture surfaces, the smoothness or ruggedness, and distortions and deformations on the crack surfaces are generally observable. In a particular type of fracture known as

---

[4] Depending on the type of fractures in question, a fracture event can take a relatively short time or a very long time. Perhaps "crack growth events" will be a more appropriate term for describing them.

[5] The terms "current" and "contemporary" mean "existing at the time of the investigation".

brittle fracture[6], the fracture surfaces often display a series of V-shaped marks known as the chevron marks. The chevron marks are informative because they point to both the fracture's origin and the direction of crack propagation.

Other properties of the fracture surfaces can only be observed using higher-resolution devices such as scanning electron microscopes[7]. Consider a type of fracture known as fatigue fracture. Unlike overload fractures caused by the application of a single force greater than the strength of the materials, fatigue fractures generally occur after many stress cycles—think about how we break a paper clip by repeatedly bending it back-and-forth with each back-and-forth being one stress cycle. Given certain types of materials[8], when the fatigue fracture surfaces are observed under the scanning electron microscope, they often show series of wave-like ridges, or furrows called "striations". Striations are microscopic features of the fatigue fracture event, and research has indicated that each striation is the result of a single stress cycle. As a result, the number of striations is informative about the number of stress cycles that have caused fatigue fracture.

The above properties of fracture surfaces are all features of the fracture event. Moreover, when the fracture surfaces continued to have these properties after the event was over, the properties could be preserved as *traces* of the fracture event, as I argue in the next section.

My second example is the encounter of an aircraft in flight with a mountain rotor, which is a type of atmospheric disturbance with a horizontal axis of rotation caused by the mountainous terrain. By an encounter, I mean the aircraft and the rotor crossing paths, and the aircraft penetrating the low-pressure core of the rotor. The temporal duration of the encounter is determined by how long the aircraft remained (wholly or partially) in

---

[6] Brittle fracture is the sudden, very rapid cracking of a part under stress where the material exhibited little or no evidence of deformation before the fracture occurs.

[7] A scanning electron microscope produces images of a sample by scanning the surface with a focused beam of electrons.

[8] It is important to note that not all engineering materials form striations during fatigue. Striations are clearly seen in pure metals and some ductile alloys. For instance, they are typically prominent in aluminum alloys. Many polymers also display well-defined fatigue striations on the fracture surfaces. However, fatigue striations occur very infrequently and are poorly defined in steels. They are often barely visible in cold-worked alloys.

the rotor's low-pressure core. The aircraft (and its occupants) and the rotor are entities involved in the encounter event.

Let us say that the temporal boundary of the encounter is $[t_1, t_2]$. Since the rotor was involved in the event, at least some of the rotor's properties during this period are features of the encounter event. For instance, the rotational speeds of the rotor during $[t_1, t_2]$ is a feature of the encounter event. More precisely, since the rotational speeds of the rotor could constantly be changing during $[t_1, t_2]$, we could say that the rotational speed of the rotor at time t_i, for any $t_i \in [t_1, t_2]$, is a feature of the encounter event. Similarly, the angle of interception (the angle at which the aircraft penetrated the low-pressure core of the rotor) is a relation between two entities involved in the encounter during the time span $[t_1, t_2]$. Thus it is also a feature of the encounter event.

Consider a trickier example of a feature of the encounter event, namely the *average* rotational speed of the rotor during the encounter. The average speed of the rotor during the encounter is obviously not a property of the rotor at any particular time during the encounter. Rather, it is a statistical property of the time history of the rotational speeds of the rotor during the time span of the event—we may say that it is a second-order property of the time history of a property of the rotor. Nevertheless, the average rotational speed of the rotor is a feature of the encounter event.

What the "average speed" example shows is that I need to expand my characterization of the first type of features. The rotor's speed is a time-indexed property, and its actual value changes from moment to moment. The rotor's average speed during the encounter is a second-order property of the time history of rotor speeds. This means that when we talk of properties (of involved entities) as features, we should include not just property instances at particular times, but also, the time history of the property instances and the second-order properties of the time history.[9] Similar examples include flight parameters:

---

[9] On the one hand, when I speak of properties as features of a particular event, I mean property tokens. For instance, the airspeed of the aircraft at a particular time is a feature of the flight event, and the specific airspeed is a property token of the property type (airspeed). On the other hand, when I discuss

A flight parameter (such as airspeed, altitude, vertical acceleration) at a particular time within a flight event is a feature of the event, but so are the time history of that parameter as well as the statistical properties of the time history.

This concludes my illustration of the first type of features of an event, namely some of the properties (and time history of the property instances) of the entities involved in the event.

The second type of feature of an event is the identities of the entities involved in the event. Metaphysically, the identities of the involved entities may be the most puzzling type of feature of events. It may be controversial what exactly are the identities, whether they can be reduced to certain properties of the entities in question, etc. I will not explore these metaphysical questions in this chapter, and I leave it open whether the first two types of features are mutually exclusive metaphysical categories.

In practice, however, the importance of the identities of those entities involved in an event cannot be denied. This is particularly clear in criminal investigations, in which the identities of the persons who committed the crimes are arguably the most important features of the crime events. Moreover, at least part of the importance is epistemic: Identifying the persons involved in a crime makes it much easier to infer additional information about how the crime event happened. In general, knowing one feature of an event can be informative about other features of the same event, and the identities of the involved entities can be a particularly informative type of feature.

It is also straightforward to find engineering examples illustrating identities as features of an event. In the TWA flight 800 case, the investigators suspected that the fuel-air explosion in the center wing fuel tank was caused by a short circuit that transferred excess voltage from a high voltage wire to a low voltage wire. The identities of the two wires are

---

feature dependence, I often transition to thinking about features as property types. Now, the dependence of a particular feature of a particular event on another feature of another event is a relationship between two property tokens. But the underlying, systematic relationships between features of event types are relationships that connect property types, which can be represented as variables.

features of the short circuit event. In the United flight 232 investigation, the investigators suspected that some rotating part of the No.2 engine flew apart, and the fragments severed the hydraulic lines in the right horizontal stabilizer. The identities of the No.2 engine parts that severed the lines—which turned out to be fragments from the stage 1 fan disk—are features of the hydraulic line severing event. Finally, in the AA flight 191 investigation, the investigators determined that something had overstressed the upper flange of the pylon aft bulkhead in a downward direction. Eventually, they discovered that the object that overstressed the upper flange was the wing clevis, which allowed them to find out that the overstress occurred during a maintenance operation.

The third and last type of feature of an event is the spatial and temporal characteristics of the event. First, the starting time point and endpoint of the event, and the time interval between them, are features of the event. For instance, in the case of a fatigue crack (created over a long time), the investigators often want to know when the crack initiated. That is the starting time of the fatigue cracking event and is a feature of the event. In the example of an aircraft encounter with a rotor, the duration of the event (the amount of time the aircraft is in or partially in the rotor's low-pressure core) is a feature of it.

Second, for a complex event E composed of smaller "subevents", the time history of the subevents (and the second-order properties of the time history) are features of the event E. For instance, the time history of the formation of microscopic striations is a feature of the fatigue crack growth event. Similarly, if a property of an involved entity constantly changes during the event, then the time history of the instances of that property (as well as the second-order properties of the time history) are features of the event. For instance, the growth rate of a crack (the derivative of the crack length relative to time) is a feature of the crack growth event.[10]

Finally, the spatial characteristics of the event—including but not limited to the location

---

[10] This is a place where type 3 features overlap with type 1 features, which is another reason why I do not claim that the different types of features are mutually exclusive categories.

of the event—are also features of the event. For instance, the location of the fracture on a piece of equipment, the location of the fracture origin, and the direction of the crack growth are all spatial features of the fracture event.

In sum, some properties of the entities involved in the event, the identities of the entities involved in the event, and the spatial and temporal characteristics of the event are three main types of features of the event. There does not appear to be an existing metaphysical category that covers all the three types of features, which is why I use a metaphysically neutral term "features" to characterize them all. In the next section, I will introduce a related notion "traces", and explain how traces can provide epistemic access to the features of some events.

8.2 Traces

As I mentioned earlier, a reverse causal inference typically starts out with a known event E and its known features, and proceeds to infer details about the causes of E. But how do investigators have epistemic access to E and its features in the first place? After all, by the time of the investigation, E had already ended, and the investigators no longer had direct access to it. One possibility is that what is known about event E was obtained via a previous application of reverse causal inference, based on features of a known effect F of E. But such a regress cannot continue forever: *Something* is needed to get the chain of reverse causal inference started. What could it be?

My answer is that *traces* serve as a major source of information for past events and their features, and that it helps to kick-start reverse causal inference. This section's main task is to develop an account of what traces are and how they provide information about features of past events.

There are three main intuitive ideas associated with the concept of traces (of past events). First, traces are current physical properties of contemporary objects that are epistemically accessible to the investigators.[11] Second, traces depend on past events in some sense: The traces were "left behind" by these past events, and some would say that traces were caused by, or were causal descendants of, the past events in question.[12] Third, in the context of reconstructing past events, traces are informative about or provide epistemic access to some of these events.[13] In the rest of this section, I will turn these intuitive ideas into a more precise account of traces, by connecting these ideas with examples of traces from engineering failure investigations. I particularly focus on the two types of commonly available traces in aviation failure investigations: Properties of fracture surfaces, and FDR (flight data recorder) data.

Fracture surfaces were created during what I call the *fracture events*, which consist of three phases: Crack initiation, crack propagation, and final separation of the parts. In the previous section, I argued that at least some current physical properties of fracture surfaces (that were created during the time span of the fracture events) are *features* of the fracture events. Now I go further and argue that those current physical properties of fracture surfaces that are *features* of the fracture event are also *traces* of the same fracture event. The argument consists of showing that all the three intuitive ideas associated with the concept of traces apply to some of the physical properties of fracture surfaces:

First, assuming that the fracture surfaces have been well preserved, their physical properties that were created during the fracture events still exist and are currently accessible to the investigators.

---

[11] Both the term "current" and the term "contemporary" refer to the time of the investigation.

[12] As we shall see, this idea will be modified in my account of traces. On my account, some traces are caused by features of past events, but other traces are *identical* with features of past events. For these second type of traces, they depend on past events in the sense that they were literally aspects of the events themselves.

[13] For instance, in chapter 3 of his book *Rock, Bone, and Ruin: An Optimist's Guide to Historical Science* ([19]), Adrian Currie distinguishes between causal accounts, informational accounts and (his own) evidential account of traces. Rather than seeing these different approaches as radically different theories of traces, however, I see them as putting different emphasis on the three intuitive ideas mentioned here; and my account will combine *all* of these intuitive ideas.

Second, (some of) the fracture surfaces' current physical properties depend on the fracture event because they were created during the fracture event, and they would not have been created if the fracture event had not occurred or had occurred differently. Some may say that the creation of the fracture surfaces and their properties was *caused* by the fracture event; I prefer to think that a fracture event *is* the creation of fracture surfaces and their physical properties.

Third, (some of) the current physical properties of fracture surfaces are informative about the fracture event, because they are aspects of the fracture event that persisted after the event was over, so of course, knowing about these properties is directly informative about the event itself.

In sum, I claim that some current physical properties of the fracture surfaces are *both* features of the fracture event and *traces* of the fracture event. These physical properties are both features and traces because they were created during the fracture event and continued to exist after the event was over.

The second type of traces commonly used in aviation accident investigations is the (Flight Data Recorder) data. The FDR records data from various aircraft sensors onto a medium designed to survive an accident. Commonly recorded parameters of the flight include altitude, airspeed, vertical acceleration, fuel flow, etc. If the data is digital, we may think of the recording medium as the object, and the recorded data (basically, a collection of symbols) as properties/states instantiated by the medium. Suppose we carve out a certain spatial and temporal part of the flight as the boundary of a "flight event". I argue that if the FDR recorded that event's parameters, then the recorded data are traces of that event. Below, I use airspeed data as a concrete example to illustrate the argument.

Airspeed is the plane's speed relative to the air around it. The pitot tube system, essentially a differential pressure gauge device, is used by airplanes to measure forward airspeed. The device measures the impact pressure of the incoming air, which equals $1/2 \times$ air density $\times$ airspeed^2. To measure the impact pressure, the device measures the pressure difference between a static sensor not in the air stream and a pitot tube in the air stream. When the airplane is at the gate, the pressure in each tube is the same and the airspeed indicator shows zero. When the airplane is in flight, the rush of air causes a pressure differential between the static tube and the pitot tube, which causes the pointer on the airspeed indicator to move. The airspeed indicator converts impact pressure to *indicated airspeed* using the equation mentioned earlier, given standard assumptions about air density. The indicated airspeed is then calibrated to account for various factors, such as instrument errors, altitude, air compressibility effects, etc. The airspeed data are then digitally stored in the FDR, and can be read out when needed.

The airspeed data recorded in the FDR satisfy the three intuitive ideas associated with the concept of traces. First, assuming that the FDR and its data survived the crash, the data can be read out and thus are epistemically available for the investigators.

Second, the airspeed data provides epistemic access to certain features of the (appropriately defined) flight event, namely the aircraft's actual airspeeds within the flight event's time span. The actual airspeed of the aircraft at a given time is a relationship between two entities—the aircraft and its surrounding air, both of which were involved in the flight event. Therefore, the actual airspeed at a given time within the flight event is a feature of the flight event, and we can have epistemic access to it based on the FDR data. The epistemic access is enabled by several equations that encode relationships between airspeed and impact pressure, properties of the atmosphere, etc.

Third, the airspeed data depends on the actual airspeed. The dependence is causal: The magnitude of the real airspeed at a given time makes a causal difference to the magnitude of the recorded airspeed at that time. Recall that in the example of fracture surfaces, the physical properties of fracture surfaces are both features of and traces of the fracture event. In that case, the traces of an event are *identical* with features of that event. In contrast, the recorded airspeed data are traces of the flight event, but they are not features of the flight event. Instead, in this case, the traces (airspeed data) of an event are *caused by* features (actual airspeed) of that event. The causal sequence can roughly be depicted as follows:

$$\text{True airspeed} \Rightarrow \text{Impact Pressure} \Rightarrow \text{Recorded airspeed.}$$

Based on the above two examples (the properties of fracture surfaces, and the FDR data), I propose the following account of traces, which is a more precise formulation of the three intuitive ideas associated with the concept of traces.

A **trace** of a past event E is a property P of an entity X, such that the following three conditions are satisfied:

1. X exists at the time of the investigation with property P, and P is epistemically accessible to the investigators.
2. P provides information or epistemic access to some feature F of event E.
3. P provides epistemic access to feature F, either because P is identical with F, or because P is causally dependent on F and the systematic dependence relationship that connects F and P is known.

Traces allow the investigators to infer the features of *some* events. In the next chapter, I discuss a type of reverse causal inference based on the idea of *feature dependence*, which allows the investigators to infer from the features of some events to the features of other events.

Chapter 9

Feature Dependence

9.1 Feature Dependence: The Basics

Feature dependence is, first and foremost, a type of *statement*, which states that features of an event depend on features of the causes of that event. Feature dependence statements can be very useful in certain types of reverse causal inference. Sometimes, I will use the term "feature dependence" or the term "feature dependence arguments" to refer to those types of reverse causal inference that make use of feature dependence statements. In this section, I introduce the basic idea of feature dependence and the basic forms of feature dependence statements. The next section addresses the evidential basis of feature dependence statements. The rest of this chapter discusses a type of feature dependence argument called "descriptive enrichment", which combines multiple features of an event to construct a description of one of its causes.

The basic idea of feature dependence can be represented as a conditional statement: "If one event C is a cause of another event E, then certain features of E depend on certain features of C." The antecedent of the conditional is a causal statement, which describes a causal relationship between two events. Moreover, in the next section, I will further argue that the consequent of the conditional is *also* a causal statement. More specifically, the consequent describes a causal relationship between two features, and the features can be represented as the values of two variables. In short, a feature dependence claim can be viewed as a conditional statement that encapsulates an inference from one causal statement to another causal statement.

Moreover, I make a distinction between two types of feature dependence statements, namely *feature necessity* and *feature sufficiency.* The basic idea of feature necessity can be characterized as: "If an event C is a cause of an event E, then, given that E has certain features, C must (or likely) have certain corresponding features."[1] That is, certain features of the outcome *require* certain features of one of its causes. In contrast, the basic idea of feature sufficiency can be characterized as: "If an event C is a cause of an event E, then, given that C has certain features, E must (or likely) have certain corresponding features." That is, certain features of a cause *imply* certain features of the outcome.

Feature necessity plays a much more prominent role in reverse causal inference than feature sufficiency, at least in the context of reconstructing past events. This is because reverse causal inference is typically used when the outcome is known while its causes are not. Feature necessity is useful when the features of the outcome are known, whereas feature sufficiency is useful when the features of the causes are known. Hence feature necessity is often a better fit for reverse causal inference than feature sufficiency, and I focus entirely on feature necessity in this chapter. Below, I provide two formulations of feature necessity, both of which will be used in later discussions of descriptive enrichment.

- Formulation 1 of feature necessity: If event E has a feature F_E, then, there must exist a cause C of E, such that C has a feature F_C.

This formulation can be used when we know nothing about the hypothetical event C yet and we need to infer its *type* first. Consider an example: "If the fracture surface created in the fracture event has chevron marks on it, then the fracture must have been caused by

---

[1] I am going to be relaxed about the use of present tense in this section, such as "C *is* a cause of E". Past tense is perhaps more appropriate given that C and E are supposed to be past events.

an overload."[2] In this example, event E is the fracture event; feature F_E is a property of fracture surfaces created in the fracture event, namely the chevron marks on the fracture surfaces; and feature F_C is the *type* of the cause C, namely an overload.

- Formulation 2 of feature necessity: If event C is a cause of event E, then, given that E has a feature F_E, C must have a feature F_C.

This formulation is often used when we already know (or assume) some partial information about the (hypothetical) event C, and we want to infer *more* information about it (given the assumption that it is a cause of the outcome E). On the surface, it may not be obvious why this formulation of feature necessity is useful for *causal inference*, given that causation is already assumed in the antecedent, and that the statement itself is more directly focused on inferring features of the hypothetical cause C. However, having more information about the putative cause C amounts to enriching a causal hypothesis, which often makes it easier to test the causal hypothesis via other methods such as additional outcomes (which I will talk about in the next chapter).

To provide more intuition about how the second formulation of feature necessity can be used in causal reasoning, consider a simple example from the investigation of the United flight 585 crash. Let event E be the loss of control of the flight, or more precisely, the continued right roll and yaw of the aircraft during a span of about 10 seconds. Let C be a hypothetical cause to be evaluated, namely a (hypothetical) encounter with a mountain rotor. The argument goes as follows:[3]

- Premise 1: If an encounter with a rotor caused the continued right roll and yaw, then, in order to account for the rate of heading change of the aircraft during the loss of control, the rotor must have followed the flight path of the accident airplane for

---

[2] An overload is an application of stress that exceeded the strength of the material.

[3] I use bullet points to represent steps in the argument, and I insert comments at various places within the argument to explain individual premises or steps.

8 seconds and increased in strength to about 1.8 radians (103 degrees) per second. ([10], p.260)

Premise 1 is an instance of formulation 2 of the feature necessity statement. Here, the feature F_E of event E is the precise rate of heading change during the loss of control. The investigators had access to this feature via a trace, namely the heading data recorded in the FDR tape. Moreover, the feature F_C of event C consists of the specific trajectory of the rotor (i.e., following the flight path for 8 seconds) and the specific strength of the rotor (1.8 radians per second) during the (hypothetical) encounter. What Premise 1 accomplishes is an enrichment of the causal hypothesis: *Assuming* that event C was a cause of event E, Premise 1 enables us to learn more about the event C, namely that C has feature F_C.

- Premise 2: However, it was extremely unlikely for there to be a rotor of 1.8 radians per second that followed the flight path of United flight 585 for 8 seconds during the accident.

Premise 2 is a rejection of the enriched causal hypothesis, in particular, the statement that there existed an event C with feature F_C. Part of the support for Premise 2 came from background knowledge: The strongest rotors ever documented in the Colorado Springs area (where the crash happened) had a strength of about 0.05 radians per second, which is much weaker than the hypothesized rotor of about 1.8 radians per second. ([10], p.260)

Another part of the support for Premise 2 came from the so-called additional outcomes argument. I will explain the abstract structure of additional outcomes argument in the next chapter, here it suffices to see the reasoning in this particular case: if United flight 585 had penetrated the low-pressure core of a rotor of 1.8 radians per second and remained there for 8 seconds, its FDR would almost certainly have recorded signature changes in indicated airspeed and altitude, and its CVR would have recorded sounds characteristic of intense rotors. None of these traces could be found in the FDR and CVR data from the accident airplane. ([10], p.260)

- Conclusion: It is very unlikely that the loss of control (the continued right roll and yaw) of United flight 585 was the result of an encounter with a mountain rotor.

The United flight 585 example illustrates how the second formulation of feature necessity was used to reach a negative causal conclusion—namely that hypothetical event C was *not* a cause of event E. Later in this chapter and the next chapter, I will also use examples to illustrate how the two formulations of feature necessity can be used to reach positive causal conclusions (i.e., that one event was a cause of another event). Before I do that, however, I want to address the following question: What is the evidential basis of feature dependence statements? For instance, what form of evidence supports Premise 1 of the United flight 585 example?

9.2 Evidential Basis of Feature Dependence

Let us reconsider the second formulation of feature necessity claim: "If event C is a cause of event E, then, given that E has a feature F_E, C must have had (or probably did have) a feature F_C". This conditional statement encapsulates an inference from a causal relationship between event C and event E, to a dependence relationship between feature F_C of C and feature F_E of E. In this section, I address two main questions: First, what is the evidential basis of such an inference? Second, what is the nature of the dependence relationship between feature F_C and feature F_E? For the first question, I propose that the inference can be supported by suitable background knowledge concerning certain *event-based causal systems*. For the second question, I propose that the dependence relationship between the two features is *also* a causal relationship.

To motivate the idea of an event-based causal system, I first introduce a distinction between statements *that* there is a causal relationship between two events, and statements about *how* one event caused the other event. To say *that* one event C is an actual cause of another event E is to make a counterfactual statement: "If C had not occurred, E would not have occurred." Different counterfactual theories of actual causation (e.g., David Lewis' or James Woodward's theory) give somewhat different characterizations of the form and the truth condition of the counterfactual conditional. But the basic idea that an actual causal claim connecting two events expresses such a counterfactual dependence is shared by these counterfactual theories of causation.

Furthermore, statements about *how* event C caused another event E can be divided into two subtypes. First, some statements can specify how C caused E by tracing through the causal pathways[4] from C to E. This first subtype of statements is used in the method of *process tracing*, which I will discuss in a later chapter.

Second, some statements can specify *how* event C caused event E by describing how various features of C make a difference to various features of E. I claim that the consequent of a feature dependence statement belongs to this second subtype. In other words, a feature dependence statement is a conditional, the antecedent of which states *that* C caused E, and the consequent of which specifies *how* C caused E (by specifying how a feature of C makes a difference to a feature of E). Moreover, the evidential basis of a feature dependence statement consists of background knowledge that bridges the gap between a statement that C caused E, and a statement about how C caused E.

What kind of background knowledge can justify the transition from *that* C caused E to *how* C caused E? My proposal is that the background knowledge obtained from prior examinations of an *event-based causal system* can play such a role.

The easiest way to understand the idea of an event-based causal system is through an example. Consider a fractured component. Its very existence tells us that a "fracture event" E had occurred.[5] Suppose that E is a fatigue fracture. To say that it is a fatigue fracture, is to say that E was caused by a cause C of a certain type—the application of cyclic loading

---

[4] We can think of a causal path from one event to another as a sequence of intermediate, causally connected events.

[5] The development of a fracture typically has three phases: crack initiation, crack propagation, and final failure and separation. Again, I understand the fracture event as encompassing all three phases, rather than just the last phase. You can think of it as a "crack growth" event if you prefer.

at a stress level less than the strength of the material. Now, *if* we know that E was caused by C, then our background knowledge in material engineering tells us that a certain causal system S—call it fatigue—was instantiated. I call the fatigue causal system *event-based*, because it is instantiated *whenever* an event of a certain type (cyclic loading at a stress level lower than the yield strength) caused another event of a certain type (crack growth).[6]

I do not have a full account of what event-based causal systems are or how they differ from other causal systems that are not event-based.[7] Instead, I propose that event-based causal systems S tend to have the following characteristics:

1. S is instantiated whenever an event of type X causes an event of type Y.
2. S supports multiple systematic dependence relationships between features of event type X and features of event type Y. To be more precise, each systematic dependence relationship here is a *type-level* relationship between one type of feature (of event type X) and another type of feature (of event type Y).
3. Each systematic dependence relationship between the two types of features is a *causal* relationship that can be characterized using the interventionist framework.[8] For instance, the relata of the causal relationships are variables, with each variable representing a type of features. Moreover, each causal relationship characterizes how a feature of event type Y would change in response to interventions on a feature of event type X, given that certain other parameters of the causal system S are held fixed.

To illustrate how an event-based causal system can support multiple causal relationships connecting features of one event (type) and features of another event (type), consider the

---

[6] We may say that the fatigue causal system is the underlying mechanism determining how cyclic loading causes crack growth. However, given the overuse of the term "mechanism" in the philosophical literature, I use the more neutral "causal system" instead.

[7] Some candidate causal systems that are not event-based may include: Dynamical systems such as the swinging of a clock pendulum; and state-based systems such as the dependence of gas pressure on gas temperature.

[8] The best known interventionist account of causation can be found in James Woodward's *Making Things Happen: A Theory of Causal Explanation* ([40]).

fatigue causal system again. I mentioned earlier that the fatigue causal system is event-based because it is instantiated whenever one type of event (cyclic loading of a certain kind) caused another type of event (crack growth). Moreover, the study of the fatigue causal system is a part of an engineering field called *fracture mechanics*. Given our current background knowledge about fracture mechanics, knowing that a fatigue causal system is instantiated gives us a lot of information about how the features of the cyclic loading make a difference to the features of crack growth. Below are two examples of type-level causal relationships between features, both of which are supported by the fatigue causal system.

The first example is a type-level causal relationship between the number of stress cycles and the number of striations on the fracture surfaces. Some materials such as aluminum alloys display well-defined fatigue striations on the fracture surfaces, with each striation corresponding to one advancement of the fatigue crack front. Moreover, research showed that given appropriate background conditions, there is a nearly one-to-one correspondence between the number of stress cycles and the number of striations on the fracture surfaces.[9] Here, the number of stress cycles is a feature of the cyclic loading event; the number of striations on the fracture surface is a feature of the fracture event; and there is a systematic relationship between the two types of features. The systematic relationship is a type-level causal relationship, because experimental interventions on the number of stress cycles would make a difference to the number of fatigue striations.

The second example is a type-level causal relationship between the stress intensity factor range and the crack propagation rate. Before introducing this example, I will first explain a fracture mechanics concept known as the stress intensity factor K. Basically, when we apply load to a part, sometimes the stress is not concentrated uniformly all over the part. When there is a crack already, stress is concentrated more on the crack tip. The stress intensity factor K represents stress intensity near the tip of a crack. The magnitude of

---

[9] A caveat here: Stage 1 of the fatigue crack growth (crack initiation) typically does not produce fatigue striations, so striations counting cannot help determine the number of cycles in this stage.

K depends on the geometry of the part, size and location of the crack, magnitude of the load on the entire part, and how the load is distributed. For instance, K is proportional to the square root of the crack length, which means that the longer the existing crack is, the greater the stress intensity is at the crack tip.

In the fatigue causal system, a fluctuating stress intensity drives the crack to grow at some rate. Let the *stress intensity range* be the difference between the maximum and minimum stresses in a stress cycle. When a stress intensity range ΔK is applied to a material for some number of cycles ΔN, this drives the crack to grow in length by a specific amount Δa. Here ΔK and ΔN are features of the cyclic loading event, Δa is a feature of the crack growth event, and there is a systematic dependence of Δa on ΔK and ΔN. Again, the systematic dependence relationship is causal.

An alternative and more standard way of understanding the systematic relationship between Δa, ΔK and ΔN is to specify how the growth rate of the crack depends on the stress intensity factor range. The growth rate of the crack length relative to the number of crack cycles is given by the ratio $\Delta a/\Delta N$. Using a continuous expression, the crack growth rate is the derivative da/dN. For a given material at a stress ratio[10], the crack growth rate da/dN and the stress intensity range ΔK can be plotted on a log-log scale. On the log-log plot, there is typically a straight-line region of da/dN over ΔK values. This region corresponds to the crack-propagation phase and is defined by the so-called Paris Equation $da/dN = A(\Delta K)^m$, where A and m are constants depending on the material in question.

In short, the Paris Equation defines a systematic relationship between the stress intensity factor range ΔK at a given time, and the crack growth rate da/dN at that time. If we think of the stress intensity factor range at a given time[11] as a feature of the cyclic loading event, and the crack growth rate at that time as a feature of the crack growth event, then the Paris Equation defines a type-level causal relationship between a feature of the cyclic

---

[10] The stress ratio R is the ratio between the minimum and the maximum stress amplitudes in the cycle.

[11] More precisely, ΔK is the stress intensity factor range in a given stress cycle. Similarly for da/dN.

loading event and a feature of the crack growth event.[12]

The examples of the striation count and the Paris Equation show that the very same event-based causal system—fatigue—can support multiple type-level causal relationships between features of one event (cyclic loading) and features of another event (crack growth). The main point is that all these type-level causal relationships among features are simply different aspects of the *same* causal system. We may say that the fatigue causal system *unifies* these causal relationships among features.

The above discussions about event-based causal systems also suggest a way of justifying feature dependence statements. Again, I use the second formulation of feature necessity for illustration: "If event C is a cause of event E, then, given that E has a feature F_E, C must (or likely) have a feature F_C." The following line of reasoning justifies the transition from the antecedent to the consequent of this conditional:

1. Suppose C is a cause of E. ("that" causation)
2. If so, our background knowledge tells us that some event-based causal system S is instantiated. ("how" causation in the second sense)
3. The event-based causal system S supports a type-level causal relationship R,[13] and knowledge about R constitutes part of our background knowledge about the event-based causal system S.
4. The dependence relationship between feature F_C and feature F_E is an instance of a type-level causal relationship R.

---

[12] The metaphysics of event causation gets a little bit tricky here. We know that the growth rate da/dN at a given time is dependent on the stress intensity factor at the crack tip at that time, and the stress intensity factor is dependent on the existing crack size, a at that time. Now, if we take cyclic loading as an entire event and crack growth as an entire event, then cyclic loading caused crack growth. However, it is possible for a feature of cyclic loading at a given time (e.g., stress intensity factor at that time) to causally depend on a feature of crack growth at an *earlier* time (existing crack length).

[13] Moreover, in order to justify *feature necessity*, the type level relationship R must have the characteristic that, for every value of the dependent variable, there is a *unique* value of the independent variable. Not all type-level causal relationships connecting features have this characteristic.

To put it even more perspicaciously, the flow of reasoning in support of a feature dependence statement is this:

> A causal relationship between two actual events ⇒ An instantiated event-based causal system ⇒ A type-level causal relationship within the system that connects features of two event types ⇒ A particular instance of the type-level causal relationship that connects features of the two actual events.

Note that the right background knowledge is required for this chain of reasoning to go through: We do not always have suitable knowledge about the event-based causal system that underlies a causal relationship between two actual events. However, if we do have suitable knowledge about the underlying causal system, we can justify a feature dependence statement by citing one known aspect of the system, namely one known causal relationship within the system that connects features of the relevant events.

However, there is a major missing piece in my account of the evidential basis of feature dependence statements so far. I have argued that the justification of a feature dependence statement relies on our background knowledge about the presence of some event-based causal system S and some type-level causal relationship R supported by S. What I have neglected to mention is that the justification of the feature dependence claim also requires an *assumption*, namely that no other causal system S^* that can interact with S and modify the causal relationship R is present.

To see why such an assumption is necessary, consider the fatigue example again. Let event C be a cyclic loading (of an appropriate sort), and event E be a (sub-critical) crack growth. Suppose we know that whenever cyclic loading causes crack growth, the fatigue causal system S is instantiated. According to some type-level causal relationship R supported by S, if event E has feature F_E, then event C must have feature F_C. For instance, F_C could be the number of stress cycles and F_E be the number of striations on the crack surface, or perhaps F_C could be the stress intensity factor range at a given time and F_E be the crack growth rate at that time. So far, the considerations support the feature necessity statement: "If event C is a cause of event E, then, given that E has a feature F_E, C must (or likely) have a feature F_C."

Now, however, suppose further that another event D—some type of corrosion—is also a cause of event E. We know that corrosion can obliterate striations, which changes the fidelity of striations count. So, if both fatigue and corrosion were causes of crack growth, the one-to-one correspondence between striation count and cycles of loading would no longer be reliable. In such a case, a feature necessity statement that infers the cycles of loading from the striation count would not be justified.

Moreover, the combination of fatigue and corrosion also modifies the Paris Equation, which defines a relationship between the stress intensity factor range and the crack growth rate. The fatigue crack growth rate is enhanced by corrosion given the same level of stress intensity range, and this effect is seen in all three phases of crack growth. The specific effects of corrosion on fatigue crack growth depend on a variety of factors, including the type of environment, fatigue load levels, and corrosion type.[14]

In short, the effect of corrosion on sub-critical crack growth can be regarded as an event-based causal system S^* in itself, and S^* interacts with the fatigue causal system S and modifies the quantitative form of the causal relationship R supported by S. If we have reason to believe that corrosion is a causal factor of the subcritical crack growth, then we are no longer justified in using the original feature necessity claim to infer features of the cyclic loading event. Instead, we should use the following, more complicated feature necessity statement: (Again, C is cyclic loading, D is corrosion, and E is crack growth.)

> If event C is a cause of event E, and if event D is another cause of event E, then,

---

[14] Common types of corrosion include pitting, exfoliation, intergranular; each will affect the crack growth rate in a particular material differently. For instance, pitting tend to be the most damaging type of corrosion. It increases crack growth rate more than any other kind of corrosion; even corrosion pits of a material's grain size can substantially increase the crack growth rate. The degree to which corrosion influences crack-growth also depends on fatigue-loads. For instance, corrosion causes a greater increase in crack-growth rates at low loads than it does at high loads.

given that event E has feature F_E, event C must (or likely) have feature F_C^*, which is different from F_C.

The evidential basis of this more complicated feature necessity statement is our background knowledge about how *two* event-based causal systems—namely, fatigue and corrosion in this case—interact with each other to influence crack growth. Of course, the statement still assumes that no *further* causal system that can modify the dependence relationship between feature F_C^* and feature F_E is at work.

In sum, part of the evidential basis of a feature dependence statement is evidence for the assumption that no other event-based causal system that can change the feature dependence relationship we are interested in is at work. If there is evidence *against* this assumption, then we need to modify the feature dependence statement, by considering how the multiple event-based causal systems at work interact with each other to influence features of a given outcome.

9.3 Descriptive Enrichment

Previously, my discussions have focused on the general forms and the evidential basis of feature dependence statements. In the rest of this chapter, I will discuss how feature dependence statements can be *used* in a type of reverse causal inference that I call "descriptive enrichment".

Suppose we know that an event E had occurred with certain features, but we do not know what had caused E. The basic idea of descriptive enrichment is that we can use *multiple* features of the known event E to construct a description of *one* of the causes of that event. It basically works as follows:

1. Given that event E has feature F_E^1, infer that it has a cause C that has feature F_C^1.

2. Given that event E also has feature F_E^2, infer that the cause C also has feature F_C^2.

3. Repeat...
4. At some point, we will be able to construct a sufficiently detailed description of C ("the cause of E that has features F_C^1, F_C^2, etc") and to check whether it actually occurred, etc.

The main task of this section is to provide a more rigorous formulation of descriptive enrichment using feature necessity statements, and to justify a crucial assumption (the so-called "one event assumption") needed for descriptive enrichment.

Earlier in this chapter, I introduced two formulations of feature necessity. The first formulation is: "If event E has a feature F_E, then, there must exist a cause C of E, such that C has a feature F_C." This formulation can be used when we know nothing about the cause yet and we need to infer its type first. The second formulation is: "If event C is a cause of event E, then, given that E has a feature F_E, C must have a feature F_C." This formulation can be used when we already have some partial information about the (hypothetical) event C, and we want to infer more information about it. My proposal is to represent the structure of a version of descriptive enrichment using a combination of these two formulations of feature necessity.

Before describing my proposal, I want to make one minor modification to my formulations of feature necessity. In my previous formulations, I used the schematic letter C to denote a hypothetical cause of event E. Consider for example the first formulation of feature necessity: "If event E has a feature F_E, then, there must exist a cause $\mathbf{C}$ of E, such that $\mathbf{C}$ has a feature F_C." The use of the schematic letter "C" in the consequent of this conditional may create the impression that we already have independent epistemic access to the event C and can refer to it directly. This impression, however, is unwarranted. All that follows from the consequent of the feature necessity claim is that *there exists an event with feature F_C*.

To prevent similar misunderstanding, when I use the two formulations of feature necessity in descriptive enrichment, I modify the formulations by replacing all uses of the schematic letter "*C*" with suitable existential quantifiers or anaphora. Below is an abstract representation of the argumentative structure of (one form of)[15] descriptive enrichment:

- Premise 1: Event E has feature F_E^1, F_E^2, etc.
- Premise 2: If event E has a feature F_E^1, then, there must exist a cause of E that has a feature F_X^1.
- Conclusion 1: There exists an event that has feature F_X^1 and is a cause of E.
- Premise 3: If **this** event that has feature F_X^1 is a cause of E, then, given that E has another feature F_E^2, **this** event must have another feature F_X^2.
- Conclusion 2: There exists an event that has features F_X^1 and F_X^2, and it is a cause of E.
- Continue the process to learn more features about **this** event, until we can run additional causal arguments (such as additional outcomes) based on what we know about *this* event.

Premise 2 is the modified version of the first formulation of feature necessity with the schematic letter "*C*" removed. Premise 3 is the modified version of the second formulation of feature necessity, with the schematic letter "*C*" replaced by the anaphora "this".

Descriptive enrichment argument requires a key assumption, namely that all the features that are inferred in the argument (e.g., F_X^1, F_X^2, etc) are possessed by *one* cause of E rather than by multiple causes of E. I call this assumption the "one event assumption". Note that the one event assumption is built into Premise 3 of the argument, because the consequent

---

[15] There are alternative forms of descriptive enrichment argument, one of which I will mention at the end of this chapter.

of Premise 3 assumes that it is *this* event with feature F_X^1, rather than some other cause of E, that has the feature F_X^2.

But what justifies the one event assumption? How do we know that one cause of E (as opposed to multiple causes of E taken together) accounts for the features F_E^1, F_E^2, etc of E? For Premise 3 of the descriptive enrichment argument, how do we know that it is *this* event with feature F_X^1, rather than some other cause of E, that has feature F_X^2?

My answers to these questions rely on the idea of an event-based causal system that I introduced in the preceding section. An event-based causal system S typically satisfies the following conditions: First, whenever a certain type of event C caused another type of event E, an event-based causal system S is instantiated. Second, the causal system S supports multiple causal regularities, each of which connects one feature of C with one feature of E. Moreover, to know the event-based causal system S, is to know *which* features of E depend on features of C, and how the dependence relationship works. Based on these ideas, my solution to the problem of justifying the one event assumption is simple: As long as we know which event-based causal system is at work, we can *choose* appropriate features of E to use in the descriptive enrichment argument, so as to ensure that they all depend on features of *one* cause of E.

Consider again Premise 3 in the descriptive enrichment argument: "If **this** event with feature F_X^1 is a cause of E, then, given that E has another feature F_E^2, **this** event must have another feature F_X^2." The one event assumption implicit in this premise can be justified in the following way: Suppose **this** event with feature F_X^1 is a cause of E. The feature F_X^1, which likely includes the *type* of **this** event, helps determining an event-based causal system S that connects **this** event and E. Our knowledge about S tells us *which* features $F_E^i, F_E^j, \ldots$ of E depend on features of *this* event, and we can choose to use *these* features $F_E^i, F_E^j, \ldots$ in the descriptive enrichment argument to infer more features of **this** event.

The above discussions about how the one-event assumption implicit in descriptive enrichment can be justified are relatively abstract. In the following, I describe a toy example of descriptive enrichment involving fatigue cracking, and comment on how the one event assumption in this example is justified.[16]

- Premise 1: Based on crack surfaces and their properties, we know for certain the existence of a fracture event E and at least some of its features (in the form of properties of crack surfaces).
- Premise 2: If event E has a feature F_E^1 (the presence of microscopic striations on crack surfaces), then, there must exist a cause of E that has a feature F_X^1 (is a type of cyclic loading below the strength of the material).
- Conclusion 1: There exists a cause of E that has a feature F_X^1, which is a *type* of cyclic loading below the strength of the material.

A comment on the argument so far: Premise 2 corresponds to the first formulation of feature necessity. In this premise, we use some feature of event E to infer the type of one cause of E. Below, I introduce two "implicit premises" about an event-based causal system that are useful for justifying the one event assumption.

- Implicit Premise A: The feature F_X^1 (cyclic loading below critical strength) and the *type* of E (crack growth), plus background knowledge, single out an event-based causal system, which we may call "fatigue".
- Implicit Premise B: The fatigue causal system connects multiple features of the event type characterized by feature F_X^1 (cyclic loading below critical strength), with multiple features of the event type "crack growth'.

A comment on Implicit Premise B: We have already discussed some of the causal regularities supported by the fatigue causal system. For example, there is a nearly one-to-one

---

[16] I want to emphasize that this is a toy example with no practical import, and the *only* point of this example is to show that knowledge about an event-based causal system can allow us to justifiably make the one event assumption in descriptive enrichment. Also, I use bullet points to represent steps in the argument, and I insert comments at various places within the argument.

correspondence between the number of stress cycles and the number of fatigue striations of the crack surface for certain types of materials. For another example, the Paris Equation $da/dN = A(\Delta K)^m$ defines a systematic relationship between the stress intensity factor range at the crack tip and the crack growth rate. Each of these causal relationships connects one type of feature of the cyclic loading event with one type of feature of the crack growth event. The fatigue causal system provides a *unification* of all these type-level causal relationships.

- Premise 3: If **this** event with feature F_X^1 (is a cyclic loading below critical strength) is a cause of E (cracking growth), then, given that E has another feature F_E^2 (a specific number of striations), **this** event must have another feature F_X^2 (a specific number of stress cycles).

A comment on Premise 3: The one event assumption implicit in this premise is justified by Implicit Premise A and Implicit Premise B. We know that the fatigue causal system at work connects some features of **this** cyclic loading event with some features of the crack growth event. One causal regularity supported by the fatigue causal system is the relationship between the number of stress cycles and the number of fatigue striations, and the number of stress cycles is a feature of cyclic loading. Therefore, we can conclude that **this** event that has feature F_X^1 (is a type of cyclic loading) also has feature F_X^2 (a specific number of stress cycles).

- Premise 4: If **this** event that has feature F_X^1 (a type of cyclic loading) and feature F_X^2 (a specific number of stress cycles) is a cause of E (cracking growth), then, given that E has another feature F_E^3 (the crack growth rate at a given time), **this** event must have another feature F_X^3 (a specific magnitude of stress intensity factor range at the crack tip at that time).

A comment on Premise 4: Since the underlying causal system is still fatigue (the reference to which is fixed by feature F_X^1 (cyclic loading) and the fact that there is a crack), the justification of the one event assumption implicit in Premise 4 is the same as in Premise 3. Here we are simply relying on another causal regularity supported by the fatigue causal system, namely the Paris Equation.

- Final Conclusion: There exists a cause of E that has feature F_X^1 (is a type of cyclic loading), feature F_X^2 (has a given number of total stress cycles) and feature F_X^3 (has a specific stress intensity factor at the crack tip at a given time).

Ultimately, the one event assumption implicit in a descriptive enrichment argument can be justified, because the underlying event-based causal system unifies multiple causal regularities, all of which connect features of *one* event with features of another event.

The fatigue cracking example discussed so far helps illustrate the structure of descriptive enrichment and how the implicit one event assumption is justified. However, it is only a toy example and has no practical import. In the next section, I describe a real example of descriptive enrichment used in the investigation of the American Airlines Flight 191 accident.

9.4 An Example of Descriptive Enrichment

To illustrate how descriptive enrichment works in practice, I describe an example of it in the investigation of American Airlines Flight 191 accident. In this accident, the airplane's left (No. 1) engine separated from the wing as the plane was lifting off from the runway. By careful examination of the wreckage, the NTSB investigators established that the No.1 engine's separation originated with a fracture of the upper flange of the pylon aft bulkhead. ([8], p.48)

A pylon is a structure designed to carry the engine, and it is connected to the wing through a few spherical bearings. Two of the spherical bearings are aligned vertically in a forward pylon bulkhead, an upright front wall attached to the forward portion to the wing. A third spherical bearing is in the aft pylon bulkhead, a smaller, upright back wall attached to a clevis fitting on the wing's underside. The pylon aft bulkhead has a series of projecting rims around its periphery; these projecting rims are called "flanges", and they help to connect the bulkhead to the pylon itself. In the Flight 191 accident, the upper side of the flange of the pylon aft bulkhead fractured, which caused a series of further fractures that ultimately resulted in the separation of the engine from the wing. ([8], p.12)

The largest part of the fracture on the upper flange of the pylon aft bulkhead was about 10 inches long. Let the event E be the fracture event that created this 10 inches long fracture. To identify what caused E, the investigators used a descriptive enrichment argument.

First, the investigators carefully examined the fracture surfaces, and found that the marks on the fracture surfaces (e.g., the presence of chevron and tear marks) were all typical of overload. ([8], p.12) We can represent this discovery as the first step of the descriptive enrichment:

- Premise 1: If event E had a feature F_E^1 (the presence of chevron and tear marks), then, there must exist a cause of E that has feature F_X^1 (is an overload—the application of stress greater than the strength of the upper flange)

Second, the investigators identified a few more properties of the fracture surfaces that were informative: The *direction* of the chevron and tear marks on the fracture indicated that the rupture progressed downward at the center of the flange, then inboard and outboard direction of the flange. The bottom portion of the fracture exhibited smearing consistent with the compression portion of a bending fracture. The smear was more prevalent—about 6 inches long—in the thinner center portion of the upper flange structure, but became less prevalent at the fracture's outer ends. ([8], p.12) Based on these features F_E^2 of the fracture event E, the investigators were able to infer another feature of the overload event, namely that it is downward bending stress applied at the upper surface of the center section of the upper flange. In short:

- Premise 2: If **this** event with feature F_X^1 (is an overload of the upper flange) was a cause of E (the 10-inch fracture on the upper flange), then, given that E had another feature F_E^2 (or more precisely, a few features including the spatial characteristics of the chevron marks, the smear on the bottom portion of the fracture, etc), **this** event must have another feature F_X^2 (is downward bending stress applied at the center section of the upper flange).
- Conclusion 1: Therefore, there existed an event with feature F_X^1 and feature F_X^2, and it was a cause of E. That is, the 10-inch long fracture was caused by the application of downward bending stress at the center section of the flange, which exceeded the flange's strength.

However, merely knowing that the cause of the fracture is a downward bending overload was still not enough. Many things could have imposed such a stress, and the event could have happened at any point during the airplane's service history. To further narrow the space of possibilities, the investigators zoomed in on another feature of the fracture event. At the center of the fracture surface, there was a crescent-shaped deformation. Moreover, the shape of the deformation exactly matched the radius of the bottom surface of the wing clevis to which the aft bulkhead was paired. This was strong evidence that the clevis was the object that imposed the downward bending stress on the center section of the upper flange. ([8], p.12-18)

To represent this step of the argument more formally, I will break it down into two smaller substeps. Let the feature F_E^3 be the specific shape of the crescent-shaped deformation, the feature F_X^3 be the specific shape of the object that applied the downward bending stress to the upper flange, and the feature F_X^4 be the identity of the object that applied the downward bending stress to the upper flange, namely the wing clevis. Here F_E^3 is a feature of the fracture event E, while F_X^3 and F_X^4 are features of the overload event. The first substep can be represented by the following premise:

- Premise 3: If **this** event with feature F_X^1 and F_X^2 was a cause of E, then, given that E had another feature F_E^3, **this** event must have another feature F_X^3.

In plain English, Premise 3 says: "If the application of downward bending stress on the upper flange was a cause of the 10-inch fracture, then, given that the fracture surface contains a crescent-shaped deformation of a specific shape, the object that applied the downward bending stress to the upper flange must have a matching shape."

The second substep is an inference from F_X^3 to F_X^4. The inference is through elimination: Among all the reasonable alternatives, the only object that has the precise shape required by feature F_X^3 is the bottom of the wing clevis, so the clevis must be the object that applied the downward bending stress (i.e., the overload must have feature F_X^4). This second substep can be represented by the next premise:

- Premise 4: If **this** event had feature F_X^3, it must also have feature F_X^4.
- Conclusion 2: Therefore, there existed an event with features $F_X^1, F_X^2, F_X^3, F_X^4$, and it was a cause of E. That is, the overstress crack in the flange was created by the bottom of the clevis applying downward bending stress to the upper flange.

Next, the investigators moved on to determine the *time* when the wing clevis had impacted the pylon aft bulkhead's upper flange. The investigators noticed that with the bulkhead to clevis attaching hardware in place, the upper surface of the flange was about 0.5 inches below the clevis's bottom. ([8], p.18) Therefore, for the clevis to have contacted the flange and deformed it, the attaching hardware through the clevis and the bulkhead's spherical bearing must have been removed. Moreover, the investigators further established that the attaching hardware remained in place throughout the accident. The attaching hardware could only have been removed when the pylon was installed or removed from the wing (i.e., during a maintenance operation). Consequently, the clevis must have contacted the upper flange during a maintenance operation. ([8], p.49)

To represent this step of the reasoning, let feature F_X^5 be the timing of the overload event, namely that it had occurred during a maintenance operation in which the clevis to bulkhead attaching hardware was removed. What this step says is that, given that the overload event consists of the clevis applying downward bending stress on the upper flange, and given that the clevis could only come into contact with the flange during a maintenance operation, it follows that the overload event occurred in a maintenance operation. More formally:

- Premise 5: If **this** event has feature $F_X^1, F_X^2, F_X^3, F_X^4$, and if **this** event was a cause of the fracture event E, then this event must also have feature F_X^5.
- Conclusion 3: Therefore, there existed an event with features $F_X^1, F_X^2, F_X^3, F_X^4, F_X^5$, and it was a cause of E. That is, the fracture event was caused by the bottom of the clevis applying downward bending stress to the upper flange during a maintenance operation in which the clevis to bulkhead attaching hardware was removed.

By this point, the investigators had completed the descriptive enrichment argument. They had arrived at a sufficiently detailed description of one cause of the fracture event, namely "the bottom of the clevis applying downward bending stress to the upper flange during a maintenance operation in which the clevis to bulkhead attaching hardware was removed". The description was sufficiently detailed because the investigators could check it against the maintenance records to see whether any actual event on record fits the description.

After examining the airplane's maintenance history documentation, the investigators found evidence that there was an actual event that fits the description they constructed. On March 29-30, about eight weeks before the crash, the accident aircraft went through a pylon removal procedure at an American Airlines facility in Tulsa, Oklahoma. The record showed that the midnight shift started the operation by removing the hardware that attached the aft bulkhead to the clevis before going off duty. Since the forward bulkhead attachment hardware was still in place, it could act as a pivot in the case of any inadvertent loss of forklift support to the engine and pylon assembly, resulting in an upward movement at the aft bulkhead's upper flange and bringing it into contact with the wing clevis. ([8], p.49)

The next day, two mechanics who performed the maintenance procedure stated that they saw the upper lug of the aft bulkhead resting against the bolts attaching the wing clevis to the wing. Basically, what they observed and reported was a spatial configuration consisting of the aft bulkhead resting against the clevis in a specific way. Post-accident tests showed that such a spatial configuration could only have obtained, if the clevis had deformed the upper flange by a 0.1 inch. Moreover, the vertical depth of the crescent-shaped deformation found on the 10-inch fracture of the airplane's aft bulkhead upper flange was exactly 0.1 inch. ([8], p.50) Tests performed by American Airlines confirmed that deformation of this magnitude would initiate an overload crack. This was strong evidence that the overload event that caused the 10-inch fracture of the upper flange indeed occurred during the maintenance operation on March 29-30, 1979.

- Final conclusion of the argument: The 10-inch overstress crack in the aft bulkhead's upper flange was created during a maintenance operation on May 29-30, 1979 at the American Airline's facility in Tulsa, Oklahoma, when the upper flange moved against the wing fitting clevis, and the clevis applied downward bending stress that deformed and cracked the upper flange surface.

This completes my discussion of the descriptive enrichment argument in the American Airlines Flight 191 example.

Before concluding this chapter, I want to point out that there are alternative forms of descriptive enrichment, whose abstract structures differ somewhat from the one I have presented earlier. For instance, the form of descriptive enrichment I have discussed uses both formulations of feature necessity statements as premises. We can, however, construct an alternative form of descriptive enrichment using *only* the second formulation of feature necessity, as follows:

- Premise 1: Event E has feature F_E^1, F_E^2, etc.
- Premise 2: If an event of type X is a cause of event E, then, given that event E has a feature F_E^1, then, **this** event of type X has a feature F_X^1.
- Conclusion 1: If an event of type X is a cause of event E, then this event has feature F_X^1.
- Premise 3: If **this** event of type X is a cause of event E,, then, given that E has another feature F_E^2, **this** event must have another feature F_X^2.
- Conclusion 2: If an event of type X is a cause of event E, then this event has feature F_X^1 and F_X^2.
- Continue the process to learn more features about **this** event of type X, *assuming* that it is a cause of E.

A key characteristic of this alternative form of descriptive enrichment is that its conclusions are always conditional statements. For instance, Conclusion 2 does not state that there exists an event of type X that is a cause of event E and has feature F_X^1 and F_X^2. Instead, it states only that if an event of type X is a cause of E, then it should have such and such features. In other words, the argument itself provides no evidence for or against the causal hypothesis that an event of type X is a cause of E; what it does is to enrich the causal hypothesis (by adding features to the hypothesized cause) *assuming* that the hypothesis is true.

The alternative form of descriptive enrichment presented above is not very useful in itself. However, when combined with other types of reverse causal inference, it can be a powerful tool for testing causal hypotheses (e.g., the claim that a hypothesized event of a certain type caused a given outcome). In the next chapter, I discuss a type of reverse causal inference called “additional outcomes”, which is often used in conjunction with feature dependence arguments such as descriptive enrichment to determine causes of past events.

Chapter 10

Additional Outcomes

10.1 Additional Outcomes: Two Versions

Reconstructing past events typically requires determining what events happened in the past, and which past events causally contributed to a given outcome. As we have seen in the previous chapter, feature dependence arguments are one type of reverse causal inference that can accomplish both tasks. In this chapter, I introduce another type of reverse causal inference that can accomplish both tasks, namely *additional outcomes*. Feature dependence and additional outcomes can be used independently. However, they can also be combined to form a larger causal argument, and we will see an example of that near the end of this chapter.

The basic idea of additional outcomes is simple. Suppose our task is to determine the causes of a known event E, and we want to know whether a certain hypothetical event C was a cause of E. We are not sure whether C had occurred or not, but we do know that, *if* C had occurred, it would have caused some *additional* outcome E^*, where E^* is a different event from the event E of interest. So, we can evaluate whether C had occurred by examining whether there is evidence (typically in the form of traces) that the additional outcome E^* had occurred. This is called *additional* outcomes because the main task at hand was to evaluate whether C caused the known event E, and E^* is a *different* event from E.

The argumentative forms of additional outcomes have a hypothetical-deductive flavor, and there are two versions of the arguments. First, there is a negative version of the additional outcomes argument, which supports the conclusion that the hypothetical event C had not occurred and therefore, could not have been a cause of the known event E. Second, there is a positive version of the argument, which supports the conclusion that the hypothetical event C had occurred, and that C was a cause of the known event E.[1] I will discuss the two versions of the additional outcomes arguments in turn.

The negative version of the additional outcomes argument can be represented as follows:

- Premise 1: If event C had occurred, there would have been some additional event E^* caused by C (or by some causal ancestor of C).
- Premise 2: Event E^* had not occurred.
- Conclusion 1: Event C had not occurred.
- Conclusion 2: It is not true that event C was a cause of the known event E.

I have two comments about this argument. First, Premise 2 is typically evaluated by looking for the traces that event E^* would have left behind. If the investigators could not find traces that E^* would have left behind, and if they could rule out the possibility that the traces had been destroyed by others means, then they could be confident that the premise was true.

Second, the idea that we can show "it is not true that C was a cause of E" by showing "C had not occurred" may be trivial, but it is very useful in practice. For example, one causal conclusion reached in the United Flight 585 investigation is the following: The flight crew's performance during the accident flight was not affected by illness, incapacitation, fatigue,

---

[1] In practice, the conclusions of the additional outcome arguments are often qualified with qualitative probabilistic claims: "It is very likely that C had occurred", or "It is likely that C was a cause of E", etc.

or other personal or professional problems. ([12], p.101) The conclusion is supported by the following premise: The flight crew did not suffer from illness, incapacitation, fatigue, or other personal or professional problems during the flight. ([12], p.45) This example follows exactly the argumentative structure: It is not true that C was a cause of E, because it is not true that C had occurred.

How did the investigators establish that C had not occurred in this example? Part of the evidence came from background knowledge: For instance, the pilots' medical and professional records showed no significant illness or professional problems. Moreover, part of the evidence came from additional outcomes: If the flight crew had experienced incapacitation, fatigue, etc. during the accident flight, they likely would have exhibited certain abnormal behaviors (e.g., failure to follow through certain routine procedures), which in turn would likely have left detectable traces (e.g., being recorded by the Cockpit Voice Recorder). The investigators, however, found no evidence of such abnormal behaviors in the CVR tape.

Let us look at another example of the negative version of the additional outcomes argument, this time from the TWA Flight 800 investigation. At one point during the investigation, the investigators established that a fuel/air explosion in the center wing tank of TWA Flight 800 had occurred (let this be the event E). They evaluated a variety of hypothetical causes of E, one of which was *auto-ignition*.

"Auto-ignition" means that a large volume of the fuel/air vapor is elevated sufficiently in temperature to the point where the vapor is ignitable without direct contact with an ignition source. Research indicates that the auto-ignition temperature at sea level for Jet A fuel (the fuel used in the accident airplane's center wing tank) is about 460 F, and it increases as altitude increases. Given this background knowledge, the investigators made the following additional outcomes argument: ([13], p.274)

- Premise 1: If auto-ignition had occurred, the amount of heat that would have been necessary to raise the temperature of the entire center wing tank to 460 F or higher would have caused significant thermal damage to the center wing tank.

- Premise 2: No evidence of such thermal damage was found on the center wing tank.
- Conclusion 1: It is very unlikely that an auto-ignition had occurred.
- Conclusion 2: It is very unlikely that an auto-ignition caused the fuel/air explosion in TWA Flight 800's center wing tank.

Premise 1 of the arguments says that auto-ignition would have caused an additional outcome E^*, which can be described as "thermal damage of the center wing tank". E^* is a *different* event from the main event E of interest (fuel/air explosion of the fuel tank), partly because the temporal boundaries of the two events do not overlap—likely, thermal damage would have occurred before the explosion. Moreover, the argument assumes that thermal damage from auto-ignition would leave traces distinct from post-accident fire damage. Therefore, if auto-ignition had occurred, and if the wreckage had been well preserved, the investigators should have been able to find thermal damage resulting from auto-ignition (which they did not).

Besides the negative version of the additional outcomes argument, there is also a positive version of the argument that concludes that a hypothetical event C had occurred, *and* that event C had been a cause of the known event E. First, the argument that the hypothetical event C had occurred is straightforward:

- Premise 1: Event E^*, an additional outcome different from event E, had occurred.
- Premise 2: E^* must (or likely) have been caused by an event C.
- Conclusion 1: Therefore, C must (or likely) have occurred.

Moreover, with suitable additional premises, the argument may further infer the following:

- Conclusion 2: C must (or likely) have been a cause of E.

The task of *formulating* the additional premises needed, and the task of *justifying* how these additional premises help to support the inference from Conclusion 1 to Conclusion 2, however, are more difficult. What I am going to do below is to propose a few *candidate premises* that I have found while examining real examples of the positive version of additional outcomes arguments. The candidate premises include:

- Candidate Premise 1: C was *capable* of causing E.
- Candidate Premise 2: (It is likely that) event E (the primary outcome) and event E^* (the additional outcome) had a common cause.
- Candidate Premise 3: For every other hypothetical event C^* that *has been considered* as a possible cause of E, there is evidence that either it had not occurred, or it was *incapable* of causing E.

I have three comments with regard to the candidate premises. First, with regard to Candidate Premise 1 ("C was capable of causing E"), I find it difficult to specify the precise sense of "capable"; all I can say is that (1) the claim is weaker than "C was a cause of E"; (2) the sense of "capable" in Candidate Premise 1 and in Candidate Premise 3 is the same; and (3) the claim will make more sense when we see an example. Actually, we will see two examples, one from the TWA Flight 800 investigation and the other from the USAir Flight 427 investigation. In both of these examples, the additional outcomes arguments use the premise "C was capable of causing E", but the precise sense in which C was capable of causing E is different in the two examples.

Second, with regard to Candidate Premise 3, note that it only says "for every other hypothetical event *that has been considered* as a possible cause of E". The premise does not guarantee that all the relevant alternative possible causes have been considered.

Third, it is not always the case that all three candidate premises are used together in an argument for Conclusion 2. In one of the two examples (the TWA 800 example) discussed below, all the three candidate premises are used. But in the other example (the USAir 427 example), only Candidate Premise 1 and 2 are used, and Candidate Premise 3 is not true in that example. I think of these candidate premises as a collection of *tools* that helps bridge the gap between Conclusion 1 and Conclusion 2, in the following sense: The candidate premises do not, individually or collectively, guarantee that Conclusion 2 is true. Rather, each candidate premise increases the likelihood that the conclusion is true.

10.2 Two Examples of the Positive Version

Finally, I turn to the two examples of the positive version of additional outcomes argument. The first example is again from the TWA 800 investigation. Recall that at one point in the investigation, the investigators knew that event E, namely the explosion of the fuel/air vapor in the center wing tank, had occurred. They wanted to determine what caused the explosion. Eventually, they concluded that the most likely cause of the explosion was a short circuit outside of the center wing tank that transferred excessive voltage to electrical wiring associated with the fuel gauge system. ([13], p.279)

The fuel gauge system, alternatively known as the fuel quantity indication system, is a sensing and indication system that allows the pilots to know how much fuel there is in each fuel tank. The system consists of electrical wires that connect fuel quantity indicators in the cockpit with probes in each fuel tank. In fact, the only electrical wiring inside the center wing tank of TWA 800 was fuel gauge wires. Because of their proximity to the fuel/air vapor, the fuel gauge wires were designed to carry very low voltage (about 25 volts) and could not discharge enough energy to ignite the fuel tank. However, the fuel gauge wires were routed in wire bundles along with high-voltage wires that powered other aircraft systems. A short circuit among these wires could have transferred higher-than-intended voltage to the fuel gauge wiring and ignited the fuel/air vapor in the tank.

To support the claim that such a short circuit had occurred and caused the explosion of the fuel/air vapor in the center wing tank, the investigators used an additional outcomes argument that can be reconstructed as follows:[2]

- **Step 1**: Most of the possible causes (potential ignition sources) of the fuel-air explosion that had been considered were ruled out, either because the characteristic traces that they would have left behind could not be found in the wreckage, or because they could not generate sufficient ignition energy to ignite the fuel/air vapor in the center wing tank.

Step 1 corresponds to Candidate Premise 3: "For every other hypothetical event C^* that have been considered as a possible cause of E, there is evidence that either it had not occurred, or it was incapable of causing E". For a hypothetical event C^* being considered, the evidence that it had not occurred consists of the lack of characteristic traces. For a hypothetical event C^* being considered, the evidence that it is incapable of causing E consists of its inability to generate sufficient ignition energy to ignite the fuel/air vapor in the tank. For instance, auto-ignition was ruled out due to the lack of characteristic traces, whereas electromagnetic interference was ruled out because it could not generate sufficient ignition energy.

- **Step 2**: A short circuit involving the fuel gauge wiring could have generated sufficient ignition energy, and the fuel gauge wires could have transferred the excess energy to the center wing tank and released it inside the tank.

Step 2 corresponds to Candidate Premise 1: "The event C was *capable* of causing event E". Here C is a short circuit that transferred excess voltage to the fuel gauge wiring. In this example, the short circuit was *capable* of causing the explosion of the fuel/air vapor in the center wing tank, in the sense that the conditions necessary for a short circuit to cause a

---

[2] I break down the argument into individual steps, and I present the argument one step at a time and insert comments after each step.

fuel/air explosion either were present, or could have been present as far as the investigators could tell.

For instance, in order for a short circuit to cause the explosion of fuel/air vapor in the center wing tank, one of the necessary conditions is that the short circuit generates sufficient energy that exceeds the minimum ignition energy of the fuel/air vapor in the center wing tank. To determine whether this condition was present in the TWA 800 accident, the investigators carried out a series of short circuit tests involving damaged wires[3] to determine the energy that could be released. The test results contained some statistical scatter, but they showed that a high voltage wire could release up to 400 mJ of energy to a lower voltage victim wire, which greatly exceeds the smallest experimentally measured ignition energy of the fuel/air vapor in the center wing tank (0.5 mJ).[4] Therefore, if a short circuit involving the fuel gauge wires had occurred in the TWA 800 accident, it *could have* produced sufficient ignition energy to ignite the fuel tank.[5] ([13], p.284)

- **Step 3**: There were multiple electrical anomalies on TWA Flight 800 right before the accident happened. These electrical anomalies included (1) two brief “dropouts” of certain harmonic tones[6] of background noises less than a second before the CVR lost power, and (2) the recovered center wing tank fuel gauge displaying a fuel quantity that disagreed by a few hundred pounds with the quantity recorded by the ground refueler.

Step 3 corresponds to Premise 1 of the positive version of additional outcomes argument:

---

[3] There was extensive evidence of pre-accident wire damage in the wreckage of the accident airplane.

[4] This minimal ignition energy was measured for the Jet A fuel used in the TWA 800 center wing tank, given assumptions about the temperature, pressure, etc. within the tank at the time of the accident.

[5] This is not the only necessary condition for a short circuit to cause a fuel tank explosion. Other necessary conditions include that the circuit breakers fail to protect against the effects of the short circuit, that there exists a mechanism that transfers the excess energy to the center wing tank and releases the energy inside the tank, etc. The investigators also examined these other necessary conditions, but I omit discussing them here to simplify the presentation.

[6] Most noises, including those generated by the electrical systems, are made up of a fundamental frequency and its multiples, which are known as harmonic tones. In the TWA 800 accident, the harmonic tones of multiples greater than 800 Hz were not recorded (“dropped out”) during the two brief moments (2 microseconds).

“Event E^*, an additional outcome, had occurred.” Here the electrical anomalies are the additional outcome E^*. Perhaps more accurately, there are multiple additional outcomes here, each corresponding to one electrical anomaly.

- **Step 4**: It is likely that one or more of these electrical anomalies were caused by a short circuit event.

Step 4 corresponds to Premise 2: “E^* was likely caused by an event C”. Again, E^* is the electrical anomalies, and C is the short circuit (involving fuel gauge wires). For instance, the cockpit voice recorder registered two dropouts in the background harmonic tones less than a second before the CVR lost power. The dropouts indicate a brief drain on the electrical power to the cockpit voice recorder, meaning that some electrical anomaly in the adjacent circuits was drawing a high amount of current at the time. This implies that event C, a hypothetical short circuit, had likely occurred. ([13], p.289)

Similarly, the recovered center wing tank fuel gauge from the cockpit displayed a reading of 640 pounds of residual fuel, which did not agree with the quantity recorded by the ground refueler (about 300 pounds). Post-accident tests done by the NTSB showed that a short circuit to the fuel gauge wiring could change the fuel gauge digital display by a few hundred pounds in less time than required to trip the circuit breaker. Therefore, the discrepancy between the fuel gauge display and the ground refueler record increases the likelihood that a short circuit involving the fuel gauge wires had occurred. ([13], p.290)

- **Step 5**: Likely, one or more of these electrical anomalies and the explosion of the fuel/air vapor in the center wing tank had a common cause.

Step 5 corresponds to Candidate Premise 2: “It is likely that the primary outcome E (the fuel/air explosion in the center wing tank) and the additional outcome E^* (the electrical anomalies) had a common cause.” The following considerations support this step: First, the dropouts of the CVR background harmonics occurred about 0.73 and 0.68 seconds before the CVR stopped recording, which means that the timing of the dropouts occurred right before the fuel tank explosion. Moreover, the fuel gauge wires connect the fuel gauge display in the cockpit and the center wing tank, which means that an entity involved in one of the electrical anomalies (i.e., the cockpit fuel gauge display discrepancy) is physically connected with an entity involved in the fuel/air explosion (i.e., the center wing tank) in a relevant way. Finally, the co-occurrence of all these anomalous events during a short time span further increases the likelihood that they had a common cause. ([13], p.289-290)

- **Conclusion**: Of the hypothetical causes of the fuel/air explosion examined by the investigation, the most likely cause was a short circuit that transferred excessive voltage to the center wing tank through electrical wiring of the fuel gauge system.

This completes my discussion of the first example of the positive version of additional outcomes argument. Next, I describe another example from the the USAir Flight 427 investigation, which also illustrates how additional outcomes arguments and feature dependence arguments can be combined together.

The accident sequence extracted from the black boxes of USAir Flight 427 showed that the aircraft crashed because of a drastic and sustained left yawing and rolling movement over a span of 24 seconds. The investigators considered a variety of possible causes of such a yawing and rolling movement, and eventually narrowed the field of hypothetical causes down to two: A mechanical rudder system anomaly called rudder reversal[7], and an anomalous flight crew action involving the pilot pressing down the wrong rudder pedal for an extended period. The investigators then focused on the following question: Which of these two hypothetical events actually caused the sustained left yaw and roll of the accident aircraft?

Strictly speaking, the investigators were choosing not between two hypothetical causes, but rather, two hypothetical causal scenarios. The rudder reversal scenario includes not just

---

[7] A rudder reversal is a movement of the rudder surface in a direction opposite to the pilot commands.

an event of rudder reversal, but also several other events and conditions, including wake turbulence[8] that initiated the entire sequence of events, a pilot command that triggered the rudder reversal, and further pilot commands in response to the rudder reversal, etc. Similarly, the pilot error scenario also includes multiple events in addition to the main event of interest (the pilot in control erroneously pressing down on the left rudder pedal until ground impact), including a wake turbulence encounter and several pilot actions leading to the fatal pilot error.

However, to simplify the presentation, I will focus mostly on the main event of interest in each causal scenario. Let the hypothetical event C be the rudder reversal, the hypothetical event D be the pilot error, the primary outcome E be the sustained left yaw and roll of USAir 427, and the additional outcome E^* be the pilot speech and breathing activities during the left yaw and roll event. The investigators ultimately concluded that C, rather than D, was an actual cause of E, and their reasoning can be reconstructed as the following positive version of the additional outcomes argument:

- **Step 1**: the NTSB's best match computer simulation of the rudder reversal scenario provided an excellent fit with the FDR data of the left yaw and roll event.

Step 1 corresponds to Candidate Premise 1 of the positive version of additional outcomes argument: "C (rudder reversal) is capable of causing E (the left yaw and roll event)". In this example, the sense in which C is *capable* of causing E is the following: C *could* have features that *could account for* certain features of E, namely the detailed history of the FDR recorded parameters during the left yaw and roll event, including heading, airspeed, acceleration, etc.

To better understand the evidential support behind this step, consider how the NTSB arrived at its best match computer simulation of the rudder reversal scenario. First, the NTSB's computer simulation model assumed the existence of certain events comprising of

---

[8] Wake turbulence is the turbulence generated by another aircraft in flight. It consists of counter-clockwise rotating vortices trailing from an aircraft's wingtips.

the rudder reversal scenario: For instance, it assumed that at some point during the accident sequence, the rudder reversed as a result of a jam of the secondary slide to the servo valve housing and moved to its left blowdown limit.[9] It also assumed an initial encounter with wake turbulence, and various pilot inputs to the flight controls during the upset event. ([10], 91-92)

Second, the NTSB derives its "best match" computer simulation of the rudder reversal scenario via an iterative computer algorithm. Very roughly, the algorithm continues to modify the features of the assumed events, until the details of the rudder reversal scenario predict flight parameter histories that are the closest match with the actual FDR data. ([10], 88) The iterative algorithm derives the following features of the events in the rudder reversal scenario: The precise timing of various events (including wake encounter, rudder reversal and pilot inputs); the severity of rudder reversal (the distance of the secondary slide jam from its neutral position); the magnitude of pilot wheel and rudder inputs, etc. We can think of the derivation of the NTSB's best match computer simulation as evidence supporting the following feature necessity statement:

> If the rudder reversal scenario had caused the sustained left yaw and roll of USAir 427, then, given that the left yaw and roll event had such and such features (e.g., the detailed history of heading, airspeed, vertical acceleration, and other FDR recorded parameters), the various events in the rudder reversal scenario must have such and such features (timing, the severity of rudder reversal, etc.).

Finally, given these derived features, the best match computer simulation produced an excellent match with the FDR data. For instance, the best match simulation resulted in a heading time history that not only matched the FDR-recorded heading within less than 1° throughout the accident sequence, but also replicated the shape of the curve that would be formed by connecting the FDR heading data points smoothly. The close fit between the

---

[9] For more details on the rudder reversal mechanism, see my case study on the USAir 427 investigation.

actual FDR data and the FDR data predicted from the best match computer simulation supports the following *feature sufficiency* statement:

> If the rudder reversal scenario had caused the sustained left yaw and roll of USAir 427, then, assuming that the various events in the rudder reversal scenario had such and such features (derived by the iterative algorithm), the left yaw and roll event would have certain predicted features (that turned out to closely match the actual features of this event).

This feature sufficiency statement does not imply that the rudder reversal scenario had caused the left yaw and roll event. However, it does capture a sense in which the rudder reversal scenario *was capable of causing* the left yaw and roll event.

In sum, we can think of the evidential reasoning in support of Step 1 of the additional outcomes argument as a complicated feature dependence argument that combines feature necessity and feature sufficiency statements.

- **Step 2**: Boeing's kinematic analysis assuming the pilot error scenario matched the FDR data as well as the NTSB's best match computer simulation of the rudder reversal scenario.

Step 2 corresponds to the statement "D (pilot error) is *also* capable of causing E (the left yaw and roll event)". Together with Step 1, it implies that there are two hypothetical causes C and D that could have occurred, both of which are capable of causing the primary event E of interest. Therefore, Candidate Premise 3 of the positive version of additional outcomes argument, which says that "for every other hypothetical event C^* that has been considered as a possible cause of E, there is evidence that either it had not occurred, or it was incapable of causing E", does not hold in this example.

The sense in which D is *capable* of causing E in this step is the same as the sense in which C is capable of causing E in Step 1. Boeing's kinematics analysis works somewhat differently from the NTSB's computer simulations.[10] Nevertheless, the general structure of the evidential reasoning in support of Step 2 is still feature dependence: If the pilot error scenario caused the left yaw and roll event, then, given the FDR recorded parameters of the left yaw and roll, we can derive features of the events in the pilot error scenario. Moreover, given the derived features of the events in the pilot error scenario, we can predict features of the left yaw and roll event that are very close to the actual features of this event.

In short, it follows from Step 1 and Step 2 that the investigators were not able to tell whether C (rudder reversal) or D (pilot error) caused E (the left yaw and roll event) *based on the features of the event E alone.* This is why the next step of the argument takes into consideration an additional outcome E^*, namely the pilot speech and breathing activities during the left yaw and roll event.

- **Step 3**: The human performance data (about the speech and breathing activities of the pilot during the upset) matched precisely with the rudder and control wheel time histories produced by the NTSB best match computer simulation, but are inconsistent with those produced by Boeing's kinematic analysis.

Step 3 corresponds to Premise 2 of the positive version of additional outcome argument, namely that "The additional outcome E^* (the speech and breathing activities of the pilot during the left roll and yaw) must have been caused by event C (rudder reversal), rather than by event D (pilot error)". Its evidential basis consists of a feature dependence argument using *feature sufficiency* claims.

Recall that in Step 1, the investigators inferred various features of the events in the rudder reversal scenario. Similarly, in Step 2, the investigators inferred various features of the events in the pilot error scenario. The main evidence for Step 3 is: Given its inferred

---

[10] Very roughly, Boeing's kinematic analysis involves curve fitting the available FDR data and deriving accelerations from these curves. The total forces on the flight control surfaces were derived from the accelerations using Newton's laws. After assuming certain events (such as wake turbulence and various pilot actions), the analysis then divides the total forces into individual contributions by the events, which in turn helps to derive specific features of these events.

features, the rudder reversal scenario predicts features of the pilot speech and breathing activities that are very close to their actual features; in contrast, given its inferred features, the pilot error scenario predicts features of the pilot speech and breathing activities that are very different from their actual features.

For instance, in about two seconds, the CVR recorded the sounds of multiple grunts and forced exhalations on the first officer's microphone channel.[11] Two specialists in breathing physiology who examined the grunts stated that they were signs of physical effort significantly greater than normal use of flight controls. One specialist concluded that the sound suggested that the first officer was struggling unusually hard as if he was experiencing an strong resistance in flight control. ([10], p.250)

When the NTSB investigators put the first officer's breathing and grunts on a timeline, they found that the grunts' timeline did not fit with Boeing's pilot error scenario at all. According to Boeing's scenario, the grunting sounds occurred after the first officer made a full right control wheel input and after a hypothesized left rudder input. However, neither of the two maneuvers would have required more than 70 pounds of force, which was relatively mild and should not cause a pilot to grunt. ([10], p.255)

In contrast, the timeline of the first officer's grunting sounds perfectly matched the NTSB's rudder reversal scenario. According to this scenario, the first grunting sound occurred about 0.4 seconds after the rudder reversal, which manifested itself in the cockpit as the rudder pedal pushing against the right foot of the first officer. This would likely cause an involuntary physical reaction by the first officer, which explained the relative softness of the grunt. ([10], p.250)

Furthermore, the CVR recorded louder grunting sounds by the first officer at about 0.6 seconds after the rudder reversal had reached hardover, and the right rudder pedal had fully pushed back with maximum displacement. According to the NTSB's analysis, the push back force on the rudder pedal at this time could reach 400 pounds, which would

---

[11] The first officer was the pilot flying in the USAir 427 accident.

explain why the first officer was grunting so loudly: He was exerting an immense effort to resist the push back pressure on the rudder pedal with no apparent effect. ([10], p.251)

In sum, because the features of the rudder reversal scenario are sufficient for the features of the pilot speech and breathing activities during the upset, whereas the features of the pilot error scenario are not, the investigators concluded that rudder reversal likely had contributed to the pilot speech and breathing activities during the upset. Therefore:

- **Conclusion 1**: Most likely, a rudder reversal had occurred.

To further conclude that the rudder reversal had caused the sustained left yaw and roll of USAir 427, the investigators relied on the following step:

- **Step 4**: The primary event (the left yaw and roll of the aircraft) and the additional outcome (the pilot speech and breathing activities during the loss of control) likely had a common cause.

Step 4 corresponds to Candidate Premise 2 of the positive version of additional outcomes argument, which states that "Event E (the primary outcome) and event E^* (the additional outcome) likely had a common cause". In this case, the investigators believed that E (the sustained left yaw and roll) and E^* (the pilot speech and breathing activities during the left yaw and roll) likely had a common cause, because these were two abnormal events whose timelines closely correlated with each other. Since the rudder reversal was likely a cause of pilot speech and breathing activities, it was likely a cause of the left yaw and roll event too.

- **Conclusion 2**: A rudder reversal most likely caused the left yaw and roll of USAir Flight 427.

The USAir 427 example illustrates how feature dependence and additional outcomes can be combined in reverse causal inference. In this example, even though the overall structure of the inference is an additional outcomes argument, most of the individual premises of the argument (e.g., Step 1, 2 and 3) are supported by feature dependence arguments. In some premises, we rely on feature necessity to infer features of a hypothetical event C, based on features of a known event E and on the assumption that C was a cause of E. In other premises, we rely on feature sufficiency and use the inferred features of C to predict features of some additional outcome E^*, and the accuracy of the predictions constitutes *tests* of the assumption that C was a cause of E.

To sum up, additional outcomes is a form of reverse causal inference that provides evidence either for, or against, the existence of a direct causal relationship between a hypothetical event and a known event. Similar to a feature dependence argument, the conclusion of an additional outcomes argument is an individual causal statement of the form "C is a cause of E" (or its negation), which typically constitutes only a small part of the narrative of the complex event we are trying to reconstruct.

In the next chapter, I discuss another causal inference method called process tracing. Unlike feature dependence and additional outcomes, process tracing is not an inherently *reverse* causal inference, and its conclusion is a *set* of individual causal statements. It is an inferential tool that complements feature dependence and additional outcomes, and is typically used when the investigators are trying to connect the dots among the subevents they have already known about.

Chapter 11

Process Tracing

11.1 Process Tracing: The Basics

In the chapter on feature dependence, I made a distinction between a claim *that* event C caused event E, and a claim about *how* event C caused event E. The claim *that* event C caused event E is typically a difference-making claim that can be captured by the counterfactual: "If C had not occurred, E would not have occurred." In contrast, a claim about *how* C caused event E either describes sequences of causally connected events that mediate between event C and event E, or describes how features of C make a difference to features of E.

Process tracing is a type of causal reasoning that concludes *that* an event C was a cause of an event E, by determining *how* event C caused event E. Specifically, process tracing determines how event C caused event E by tracing *sequences* of causally connected events that mediate between event C and event E. I emphasize the word "sequences" because process tracing may trace more than one causal sequence that connects C and E; we may say that in general, process tracing identifies a *directed causal graph* that connects event C and event E.

Even though process tracing is a useful method for reconstructing complex events of the past, it is not really a "reverse" causal inference method, for two reasons: First, often the events C and E are both known; even some of the mediating events between C and E may have been known already. A major part of process tracing is to "connect the dots" by establishing causal connections among these known events and adding suitable hypothetical events and conditions when necessary. Second, to connect event C and event E, process tracing could trace forward from event C, or trace backward from event E, or do both until reaching some middle point. In other words, the procedure of process tracing is flexible and does not have a fixed direction (e.g., from effects to causes).

Another characteristic that distinguishes process tracing from reverse causal inference methods such as feature dependence and additional outcomes is that, instead of concluding with a single causal statement stating a direct causal relationship between event C and event E, process tracing concludes with a *set* of causal statements, each of which describes an individual link in a causal "chain" that connects C and E. Given this characteristic, process tracing is typically used when it is easier or more feasible to establish these intermediate causal connections than to establish a direct causal connection between C and E. Sometimes the intermediate causal connections are easier to establish, because they can be subsumed under well-defined type-level causal claims, whereas a direct causal relationship between event C and event E cannot.

In this chapter, I describe *a* systematic way of carrying out process tracing, which involves repeated applications of J. L. Mackie's concept of the INUS conditions.[1]

The term "INUS condition" stands for a "insufficient but necessary part of an unnecessary but sufficient condition".[2] For instance, consider an expression $(A \wedge B \wedge C) \vee (D \wedge E \wedge F) \Leftrightarrow G$, where $\wedge$ means "and", $\vee$ means "or", and $\Leftrightarrow$ means "necessary and sufficient conditions". Here the schematic letters A, B, C, etc. represent *type level events and conditions*, although sometimes I also use the letters to represent *instances* of the relevant

---

[1] I do not claim that this is the *only* way to carry out process tracing, either in engineering failure investigations or in other types of event reconstructions.

[2] For an account of INUS conditions, see Mackie, J. L., 1965. "Causes and Conditions". [27].

types for convenience. Similarly, the expression $(A \wedge B \wedge C) \Rightarrow G$ represents a type-level regularity, which says that event type A and event type B and event type C together are *causally sufficient* for event type G.[3] I call the set $\{A, B, C\}$ a *sufficiency set* for G, since it is a set of events and conditions that, taken together, is causally sufficient for G. Similarly, $\{D, E, F\}$ is a sufficiency set for G. Any non-redundant member of a sufficiency set for G is an INUS condition of G.

It is worth noting that the type level causal sufficiency and causal necessity claims that underlie the notion of INUS conditions are *ceteris paribus claims*: They are not universal, and there are situations in which they do not hold.[4] One way of interpreting the ceteris paribus status of a causal sufficiency claim is to think of it as being defeasible by the presence of certain disturbing factors: For instance, the type level generalization $(A \wedge B \wedge C) \Rightarrow G$ says that event or state types A, B, C together are typically sufficient to cause another event type G, provided that disturbing factors are absent. Similarly, causal necessity claims such as $(A \wedge B \wedge C) \vee (D \wedge E \wedge F) \Leftarrow G$ are also defeasible by the presence of disturbing factors: The causal options that are *typically* necessary for G are not *always* necessary, because there are exceptional circumstances where highly unusual causes produce G.[5]

INUS conditions are useful for process tracing for a few reasons. First, process tracing involves both tracing forward from an event and tracing backward from an event. In order to trace forward from an event C, the concept of *sufficient condition* is often needed: We need to know what events that C, together with other events and conditions, is sufficient

---

[3] Unlike Mackie, I will interpret the notions "necessary conditions" and "sufficient conditions" as *causal*. This works for me because my goal is not to give a reductive account of causation using INUS conditions, but rather, to show how INUS conditions can be used in a particular type of causal inference (process tracing)

[4] In [27], Mackie introduces the notion of a *causal field* to capture such a lack of universality. A causal field is the context or background conditions against which a causal statement is evaluated. A causal regularity that holds in one causal field may not hold in another causal field. I will use the notion "ceteris paribus" instead, as it is more standard in the literature on regularities and laws. See Alexander Reutlinger, Gerhard Schurz and Andreas Hüttemann, "Ceteris Paribus Laws" ([31]), for a survey of philosophical accounts of ceteris paribus generalizations.

[5] It is often difficult, if not impossible, to explicitly list all the possible disturbing factors that could defeat a causal sufficiency or causal necessity claim. Nevertheless, for many of such claims used in engineering failure investigations, the analysts can often recognize, given a particular situation, whether there are disturbing factors that defeat the causal generalization in question or not.

for. In order to trace backward, the concept of *necessary condition* is often needed: We need to know what events and conditions are necessary for E. INUS conditions combine the notion of sufficient conditions and the notion of necessary conditions, since they are necessary conditions *relative to* a sufficiency set. Hence INUS conditions are useful both for tracing forward and for tracing backward.

Second, process tracing requires using type-level causal generalizations for establishing the intermediate causal connections, and the claims about the regularities that underlie INUS conditions are type-level causal generalizations.

Third, INUS conditions also capture the idea that a complex set of conditions must be satisfied for causation to happen. When we consistently use INUS conditions to carry out process tracing, we can identify not just one causal path between two events C and E, but (in principle) *all* the causal paths between C and E. This is useful in practice partly because it provides a more complete picture of how C contributed to E, and partly because it provides more opportunities for interventions (e.g., more ways of severing the causal relationship between C and E in future cases).

How do we use the concept of INUS conditions to carry out process tracing? I propose that we construct *sufficiency sets* when we trace forward or backward from an event to connect it causally with another event. First, suppose we are tracing forward from an event X to connect it with a downstream event E (the target event in this case). The procedure of tracing forward can be described as follows:

- We examine the other *known* events and background conditions that were more or less concurrent (i.e., occurring simultaneously) with X, and determine which of them can be combined with X to form a sufficiency set S that implies some other event Y.
- The choice of S and Y is guided by the following heuristics: Since the goal is to connect X and the target event E eventually, we want the intermediate event Y to be (prima facie) causally relevant to E. Moreover, while constructing the sufficiency

set S for Y, we ensure that S does not have any redundant members, so that every member of S is an INUS condition of Y.

- The above heuristics are used to determine what S and Y are. Once we identify these, the justification that members of S indeed caused Y will depend partly on our confidence in the causal regularity that underlies the sufficiency claim $S \Rightarrow E$, and partly on our confidence in the application of the sufficiency claim to this particular case—for instance, that the ceteris paribus clauses of the sufficiency claim are not undermined in this particular case.
- We then proceed to connect Y and E, either by tracing forward from Y, or by tracing backward from E, depending on which option is more feasible.

Second, suppose we are tracing backward from a given event Y to connect it with an upstream event C (the target event in this case). We want to identify a sufficiency set S for Y, such that every member of the sufficiency set S (1) actually occurred (or at least, we have no reason to doubt that it occurred), and (2) is an INUS condition for Y. The following is a description of the procedure that investigators often go through to construct the sufficiency set S from scratch:

- We started out with $S = \{\}$.
- We examine *known* events and background conditions that occurred during the time span between C and Y, and identify all the *known* events or conditions $X_1, \ldots, X_i$ that were *prima facie causally relevant*[6] to Y. We add these events and conditions to S.

---

[6] By "primary facie causally relevant to Y" I mean: "is a necessary member of *some* sufficiency set for Y". For instance, given a causal generalization $(V \wedge W \wedge X) \Rightarrow Y$ and the fact that X and Y actually occurred, we can say that X is prima facie causally relevant to Y, without knowing whether V and W actually occurred or not, and without knowing whether the generalization $(V \wedge W \wedge X) \Rightarrow Y$ holds in this particular case or not. Therefore, "X is prima facie causally relevant to Y" is weaker than "X is a cause of Y".

- Certain known background conditions or events, and certain features of the event Y (that we are tracing backward from) can help us infer the existence of certain hypothetical events or conditions that were *prima facie causally relevant* to Y. If there is no counter-evidence that casts doubts on the existence of these hypothetical events, we add them to S as well.[7]
- If a partially constructed sufficiency set $S = \{X_1, \ldots, X_i\}$ is not yet sufficient for Y, then the current members $X_1, \ldots, X_i$ may provide us further information about what other causes should (most likely) be included in S.
- We continue to add events and conditions to the set S until it becomes a sufficiency set for Y.
- After constructing the sufficiency set S for Y, we then determine which members of the set are redundant and eliminate the redundancies. The non-redundant members are INUS conditions—and causes—of the event Y.
- Finally, we proceed to find causal connections between C on the one hand, and the INUS conditions in the sufficiency set S on the other hand. Again, we can do this either by tracing forward or tracing backward. When C is an abnormal event, we will be particularly interested in connecting abnormal events in the set S with C.

The above procedures of tracing forward and tracing backward are primarily *discovery* procedures. Given a pair of events C and E, the investigators apply these procedures repeatedly to discover the causal graph connecting C and E. The *justification of* the results of these procedures is more complicated and requires a form of evidence that I call *coherence.*

To get an intuitive sense of what coherence is and why it is relevant to the justification of an instance of process tracing, recall that the conclusion of process tracing is not a single

---

[7] Since the goal of this step is to identify hypothetical causes of Y, we can make use of reverse causal inference methods that we discussed earlier, such as feature dependence and additional outcomes.

causal claim, but rather, a *set* Γ of causal claims. If we use INUS conditions to carry out process tracing, then each member γ of the set Γ can be subsumed under a type-level causal sufficiency claim "$\{X_1, \ldots, X_n\}$ is a sufficiency set for Y." As a result, the total evidence for a particular application of process tracing has three dimensions:

First, for each individual causal claim $\gamma \in \Gamma$, there is a question about how much *type level* evidence we have for the type-level causal generalization $(X_1 \wedge \ldots \wedge X_n) \Rightarrow Y$ that subsumes γ. For instance, we may want to know whether the type level generalization $(X_1 \wedge \ldots \wedge X_n) \Rightarrow Y$ can be derived from an accepted theory, or established through controlled experiments.

Second, for each individual causal claim $\gamma \in \Gamma$, there is a question about how we know that the type level generalization $(X_1 \wedge \ldots \wedge X_n) \Rightarrow Y$ that subsumes γ *applies in this particular case.* For instance, we may want to look for evidence that $X_1, \ldots, X_n$ actually occurred in this case, and that no disturbing factor that could undermine the generalization was present in this case.

Third, there is the question about how well different causal claims in the set "fit together", and how well they—taken together—account for the totality of all the available traces. I call this dimension of evidence "coherence", and I will discuss it more fully in the next part of the dissertation.

This completes my abstract discussions of process tracing. In the next section, I discuss a concrete example of process tracing using INUS conditions, taken from the American Airlines Flight 191 investigation.

11.2 An Example of Process Tracing

Recall that in the American Airlines Flight 191 accident, the left (No. 1) engine and pylon assembly of the DC-10 airplane separated from the wing when the plane was lifting off from the runway at Chicago O'Hare International Airport. This extraordinary event gave rise to the two major questions that defined the investigation of this accident: First, why did the engine and pylon assembly fall off? Second, Given that the engine and pylon assembly fell off, why did the aircraft crash? In the chapter on feature dependence, I described how the investigators used a descriptive enrichment argument to answer the first question. In this section, I describe how the investigators used process tracing to answer the second question.

Let event **C** be **the separation of the No.1 engine and pylon assembly from the wing** (abbreviated description: engine separation), and let event **E** be **the ground impact of the aircraft** (abbreviated description: crash).[8] The investigators knew that both events had occurred, and they wanted to know whether C had causally contributed to E in any way, and if so, how.

Moreover, based on the data retrieved from the black boxes (specifically, the FDR), the investigators also knew at least some of the events that happened within the time span between C and E. The engine separation happened right at the takeoff. The plane normally climbed for the first 9 seconds of the flight, during which it accelerated to 172 knots. For the next 11 seconds, the plane slowed down to 159 knots. At that point, it began to roll to the left, and the nose began to drop. The FDR recorded right rudder, right wing down aileron and up elevator inputs from the pilots after the initiation of the left roll, but the left roll and nose drop continued for the next 11 seconds. Three seconds before impact, the plane was banking 90 degrees to the left and perpendicular to the ground. By the time it crashed, it had rolled over onto its back and was diving nose-first to the ground at 21 degrees. ([8], p.5)

- **Step 1**: Identifying the sufficiency set for the sustained left roll and nose drop during the final 11 seconds of the flight, which includes the stall of the left wing and the lack of timely stall recovery by the pilots.[9]

---

[8] Because I discuss a large number of (sub)events of the failure sequence in this example, I use a separate alphabetical letter to represent each subevent, and use boldface to highlight the description of the event when it is first introduced. To ensure that the reader remembers what each letter refers to when mentioned again, I often add an abbreviated description to it as a reminder.

[9] Given the complexity of the example, I choose to break down the entire process tracing procedure that

Let event **D** be the **sustained left roll and nose drop during the final 11 seconds of the flight** (abbreviated description: left roll). Given that the left roll and nose drop was severe enough to make the airplane flip onto its back, it was presumably sufficient to cause the airplane's crash. We can represent this causal relationship as:

$$D \Rightarrow E$$

Where $\Rightarrow$ means "is causally sufficient for", and D and E stand for two actual events, corresponding to the abbreviated descriptions "left roll" and "crash" respectively.

However, representing the type-level causal regularity that subsumes this actual causal relationship is not straightforward, and we need to choose the descriptions of the event types in question carefully. For instance, not all the left rolls and nose drops of an aircraft result in crashes. The rate and the duration of the roll, and the altitude at which the roll initiates all make a difference to whether a crash ensues. To simplify the discussion, I will omit articulating explicitly the type level causal regularity that subsumes the actual causal relationship $D \Rightarrow E$. However, the reader should keep in mind that a proper formulation of the type-level causal regularity requires some work.

But what had caused event D (left roll) in the first place? To answer this question, the investigators traced backward from D. By examining known events and conditions concurrent with or before event D, the investigators found that the pilots' flight control inputs during the last 11 seconds were informative about the causes of D. According to the FDR data, the pilots were trying to force the right wing down and nose up, by applying the right rudder, right-wing-down aileron deflections, and nose-up elevator deflection. ([8], p.5) Under normal circumstances, these inputs should have corrected the left roll and nose drop of the aircraft, but not in this particular accident. Whatever caused D had overcome the influence of the pilots' flight control inputs and forced the aircraft to continue the left

---

the investigators went through into five steps. The division into these steps is somewhat arbitrary, and the main point of it is to make it easier for the reader to keep track of the overall structure of the reasoning process.

roll and nose drop.

Given the investigators' background knowledge about aerodynamics, the most likely candidate that can overcome the pilot inputs and cause such a continued left roll and nose drop of the aircraft was **a stall of the left wing** (call it **G**, abbreviated description "stall"), i.e., the left wing could not produce enough lift to sustain the flight. In other words, it is very likely that G (stall) is a member of the sufficiency set for D (left roll).

However, G by itself is not sufficient for D, which is *sustained (and severe)* left roll and nose drop for 11 seconds. This is because the accident aircraft had stall warning systems, and the pilots had been trained to respond quickly to stall and to follow standard stall recovery procedures. These procedures included pointing the aircraft's nose further down to reduce the angle of attack and add power to the engines to increase the speed. Post-accident tests showed that if the pilots had followed these procedures at the onset of the stall, the aircraft could have recovered, and the continued left roll could have been averted.

The FDR data showed that in this accident, **the pilots did not follow the standard stall recovery procedures at the onset of the stall** (call this event **H**, abbreviated description "no stall recovery"). Given the stall of the left wing and the lack of timely stall recovery, the aircraft was guaranteed to continue its left roll and nose drop. In other words, the set $\{G \text{ (stall)}, H \text{ (no stall recovery)}\}$ formed a sufficiency set for D (left roll), which in turn was sufficient for E (crash). We can represent this as follows:

$$\{G, H\} \Rightarrow D \Rightarrow E$$

Where $\Rightarrow$ again means " is causally sufficient for". Within the sufficiency set $\{G, H\}$, Both G (stall) and H (no stall recovery) are INUS conditions for D (left roll) because neither is redundant. Since both G and H are abnormal events, the investigators raised the following question: Is the event C (engine separation) causally relevant to the stall of the left wing (event G) and the lack of standard stall recovery response by the pilots (event H)?

- **Step 2**: Identifying a causal path from the engine-pylon separation to the stall of the left wing, which goes through the retraction of the left leading edge slats.

To determine whether C (engine separation) caused G (stall), the investigators traced forward from C. Examination of the wreckage of the left wing's leading edge showed that when the engine-pylon assembly separated from the aircraft, it severed four hydraulic lines routed through the leading edge. These hydraulic lines were the operating lines for extending and retracting the (outboard) leading-edge slats. ([8], p.11) In short, event C (plus background conditions such as the location of the hydraulic lines relative to the engine-pylon assembly on a DC-10) was sufficient to cause event **K** (**severing of the hydraulic lines connected to the left wing leading edge slats**, abbreviated description "hydraulic lines severing"). To simplify the presentation, I ignore the background conditions and represent the causal sufficiency relationship as follows:

$$C \Rightarrow K$$

The investigators continued to trace forward from event K (hydraulic lines severing) using their background knowledge. They knew that during a normal takeoff, the leading edge slats are extended to provide an increased aerodynamic lift on the wings. On a DC-10, when the slats are extended, hydraulic fluid is trapped in the actuating cylinder and operating lines. The incompressibility of the hydraulic fluid is the only thing in the design that locks the slats in the extended position. Therefore, when the hydraulic operating lines for the leading edge slats were severed, the trapped hydraulic fluid was lost, and air loads forced the left leading edge slats to retract. ([8], p.54)

The reasoning in the preceding paragraph can be captured as a sufficiency set: Let event K be the severing the hydraulic lines that operate the left wing leading edge slats, condition **J** be **the impact of incoming air** (abbreviated description "air pressure"), and condition **I** be **the design that provides no way of locking the slats in extended positions other than the hydraulic pressure** (abbreviated description "no slat locking"). Taken together, I, J, and K were causally sufficient for event **L**, which is **the retraction of the left wing leading edge slats** (abbreviated description "slat retraction"). We can represent this as

$$\{I, J, K\} \Rightarrow L$$

Which says that I (no slat locking), J (air pressure) and K (hydraulic lines severing) together formed a sufficiency set for L (slat retraction).[10] Moreover, I, J and K are all INUS conditions of L. In particular, condition I, the design that did not provide any way of locking the slats in extended positions other than hydraulic fluid trapped in the lines, caught the attention of the NTSB investigators. The NTSB ended up recommending a redesign of the leading edge slat system to include a mechanical locking device, which amounted to the elimination of condition I. Implementing this safety recommendation made it much less likely that severed hydraulic lines to the leading edge slats would cause slat retraction during critical phases of the flight.

The investigators further traced forward from event L (slat retraction). Based on background knowledge about aerodynamics, they inferred that when the left (outboard) leading-edge slats retracted, the lift on the left wing was reduced, and the airspeed at which that wing would stall was increased. Given the configurations of the accident aircraft, the retraction of the slats (event L) was sufficient to cause **the increase of the stall speed for the left wing to 159 knots** (call this event **M**, abbreviated description "stall speed increase"), which is causally relevant to G (stall of the left wing). ([8], p.54) Omitting from the sufficiency set background conditions specifying the configurations (e.g., general aerodynamic characteristics of the DC-10) of the accident aircraft, we can represent this sufficiency relationship as:

$$L \Rightarrow M$$

---

[10] Again, we simplify the representation of the causal sufficiency relationship by ignoring some background conditions, such as the fact that the leading-edge slats are normally extended during takeoffs.

In sum, by tracing forward from event C (engine separation), the investigators were able to infer multiple causal sufficiency statements. Using commas to separate different causal sufficiency statements, we can represent the results of the forward tracing as follows:

$$C \Rightarrow K,\ \{I, J, K\} \Rightarrow L,\ L \Rightarrow M, \text{ and } M \text{ is part of a sufficient condition for } G.\ ^{11}$$

Note that these results contained the *first* causal path from C (engine separation) to G (stall).

- **Step 3**: Identifying a second causal path from the engine-pylon separation to the stall of the left wing, which goes through the deceleration of the aircraft.

Even though the increase of the stall speed for the left wing to 159 knots (event M) was in the sufficiency set for the stall of the left wing (event G), M by itself was not sufficient for G. However, based on the FDR data, the investigators also knew another fact, namely that **the aircraft was flying at a speed of 159 knots about 20 seconds into the flight** (call this condition **N**, abbreviated description "flying at increased stall speed"). Moreover, the combination of M and N was sufficient for G, the stall of the left wing. We can represent this causal sufficiency relationship as:

$$\{M, N\} \Rightarrow G$$

Moreover, the FDR data also allowed the investigators to trace backward from condition N (flying at increased stall speed). After taking off, the aircraft flew normally for the first 9 seconds, accelerating to a speed of 172 knots. However, it then began to decelerate, and it only stalled after 11 seconds of deceleration, by which time it had an airspeed of 159 knots, which happened to be the new stall speed for the left wing. As a result, the left wing stalled at about 20 seconds into the flight. ([8], p.5)

---

[11] Here is a reminder about what the letters all mean: C (engine separation), I (no slat locking), J (air pressure), K (hydraulic line severing), L (slat retraction), M (stall speed increase), G (stall).

In other words, it was clear from the FDR records that N (flying at stall speed) was caused by a known event **O** (**pilots decelerated the aircraft from 172 knots for 11 seconds before the stall**, abbreviated description "pilot deceleration"). For simplicity, we can represent this as a sufficiency relationship:

$$O \Rightarrow N$$

Event O, however, is an abnormal event, because normally an aircraft accelerates during takeoff to gain lift. This gave rise to a question: Was event C (engine separation)—which is also an abnormal event—causally relevant to event O (pilot deceleration)? If so, how?

To determine if there was a causal connection between C (engine separation) and O (pilot deceleration), the investigators again traced forward from C. After examining the wreckage of the electrical system that connects the cockpit and the left wing, the investigators found that the pylon-engine separation did not disable the power meter for the No.1 (left) engine or the "engine out" warning light (which indicates the loss of power of the engine). This means that the engine separation (event C) was sufficient to cause **engine failure warnings in the cockpit** (call this event **P**, abbreviated description "engine failure warnings"). That is:

$$C \Rightarrow P$$

The investigators regarded event P (engine failure warnings) as causally relevant to O (pilot deceleration), because they were able to identify a few events and conditions concurrent with P that, when combined with P, were sufficient to cause O. First, **according to American Airlines' emergency engine failure procedure, an aircraft with an engine failure during takeoff should climb out at V2 speed until clearing a certain altitude.** ([8], p.45) Call this condition **Q** (abbreviated description "emergency procedure"). The V-2 speed is called the takeoff safety speed, which was calculated to be 152 knots for the accident aircraft prior to the takeoff.

Second, **about 9 seconds into the flight, American Airlines flight 191 was flying at a speed of 172 knots, which was higher than its pre-calculated V2 speed of 152 knots**. Call this condition **R** (abbreviated description "high initial speed").

Third, **the pilots did not realize that the stall speed for the left wing had increased, and that the increased stall speed (159 knots) was higher than the pre-calculated V2 speed (152 knots).** Call this condition **S** (abbreviated description "ignorance of stall speed increase"). Note that the condition S was *hypothetical*: Because the Cockpit Voice Recorder lost power immediately after engine pylon separation, there was no pilot conversation on record afterward, and the investigators had no direct access to what the pilots knew or did not know during the flight. Nevertheless, S (ignorance of stall speed increase) was needed alongside P (engine failure warnings), Q (emergency procedure), R (initial high speed) to form a sufficiency set for O (pilot deceleration).[12] Hence the investigators added S to the sufficiency set for O as well.

Given a few other reasonable assumptions that we have no reason to doubt in this case—such as the assumption that the pilots would follow the emergency procedure unless they had compelling reasons not to, the events and conditions P, Q, R and S together are causally sufficient for the event O (pilot deceleration). That is:

$$\{P, Q, R, S\} \Rightarrow O$$

This sufficiency relation can be described more fully with the following narrative: After seeing the engine failure warnings in the cockpit (P), the pilots consulted the American Airlines' engine failure emergency procedure (Q). Since the aircraft was flying at a speed (172 knots) higher than the V2 speed (152 knots) at the time, the emergency procedure instructed the pilots to decelerate towards the V2 speed (R). Moreover, the pilots did not realize that the stall speed for the left wing had increased to 159 knots (S), so they did

---

[12] This is because, if the pilots had known that the stall speed had increased to a value higher than the V2 speed, they would not have decelerated towards the V2 speed even if the emergency procedure asked them to.

not know it was dangerous to decelerate towards V2 speed. Since they had no reason not to follow the emergency procedure, the pilots decelerated the aircraft during the next 11 seconds (O).

In sum, the results of process tracing in Step 3 can be represented as follows:

$$C \Rightarrow P,\ \{P, Q, R, S\} \Rightarrow O,\ O \Rightarrow N, \text{ and } \{M, N\} \Rightarrow G.$$ [13]

Note that these results helped the investigators to identify a *second* causal path between C and G.

- **Step 4**: Identifying a third causal path from the engine-pylon separation to the stall of the left wing, which goes through the disabling of the slat position indication system and the slat disagreement warning system.

Next, the investigators were able to identify a *third* causal connection between C (engine separation) and E (stall) by focusing on the hypothetical[14] condition S, namely that the pilots did not realize that the stall speed for the left wing had increased (to a value greater than the pre-calculated V2 speed). They raised the following question: Why did the pilots not realize that the stall speed for the left wing had increased? After all, the accident aircraft was equipped with the slat position indication system and the slat disagreement warning system, both of which (if functioning) should have informed the pilots about the retraction of the left leading edge slats. Knowing that the left leading-edge slats had retracted should have given the pilots pauses about decelerating the aircraft, even if they were unable to quickly calculate the numerical effects of the retraction on stall speed.

To answer the question about why condition S (ignorance about stall speed increase) had obtained, the investigators traced forward from the event C (engine separation) to determine

---

[13] For a reminder of what the letters stand for, we have: C (engine separation), P (engine failure warnings), Q (emergency procedure), R (high initial speed), S (ignorance of stall speed increase), O (pilot deceleration), M (stall speed increase), N (flying at increased stall speed), G (stall).

[14] I mentioned earlier that S was hypothetical, since the investigators had no access to what the pilots knew or did not know during the entire flight, due to the failure of CVR immediately after the engine-pylon separation.

if there was a causal connection between C and S, and they found one. By examining the electrical systems in the wreckage, the investigators discovered that the separation of the engine-pylon assembly damaged certain electrical wire bundles, which caused some (but not all[15]) flight instruments in the cockpit to lose power. The instruments that lost power included the (left) slat position indication system and the slat disagreement warning system. ([8], p.44) Let us say that event C (engine separation) was sufficient to cause event **T**, which is that **the slat position indication system and the slat disagreement warning system in the cockpit were disabled** (abbreviated description "slat warning systems disabled"). We can represent this as:

$$C \Rightarrow T.$$

Furthermore, event T, together with the condition that **the wing could not be seen from the cockpit** (condition **U**, abbreviated "wing not visible"), guaranteed that **the flightcrew did not know about the left leading edge slat retraction** (condition **V**, abbreviated "ignorance of slat retraction"), which in turn guaranteed that the flight crew did not know that the stall speed of the left wing had increased (condition S):

$$\{T, U\} \Rightarrow V,\ V \Rightarrow S$$

In sum, the third causal path from the engine separation (C) to the stall of the left wing (G) can be found in the following collection of sufficiency statements:

$$C \Rightarrow T,\ \{T, U\} \Rightarrow V,\ V \Rightarrow S,\ \{P, Q, R, S\} \Rightarrow O,\ O \Rightarrow N, \text{ and } \{M, N\} \Rightarrow G.$$ [16]

By this point, the investigators had completed process tracing between C (engine separation) and G (stall of the left wing) and had identified three causal paths between the two

---

[15] For instance, we mentioned earlier that the left engine failure warning systems in the cockpit did not lose power due to the engine-pylon separation.

[16] The letters represent: C (engine separation), T (slat warning systems disabled), U (wing not visible), V (ignorance of slat retraction), P (engine failure warnings), Q (emergency procedure), R (high initial speed), S (ignorance of stall speed increase), O (pilot deceleration), M (stall speed increase), N (flying at increased stall speed), G (stall).

events. However, they had not completed process tracing between C (engine separation) and E (crash). Recall that $\{G, H\} \Rightarrow D \Rightarrow E$, where G is the stall of the left wing, H is the condition that pilots did not follow standard stall recovery procedure, D is the continued left roll and nose drop, and E is the crash. To complete process tracing between C and E, the investigators proceeded to examine whether there existed causal pathways between C and H.

- **Step 5**: Identifying a causal path from the engine-pylon separation to the lack of stall recovery response by the pilots, which goes through the disabling of the stall warning system.

Why did the investigators want to examine whether there existed causal paths from C (engine separation) to H (no stall recovery)? This is because like C, H is also an anomalous event. After all, the accident aircraft had stall warning systems that provide warnings when the aircraft approaches stall, and the pilots had been trained to respond to the onset of stall warnings by immediately carrying out standard stall recovery procedure (pointing the aircraft nose down and accelerating the engines). Given these conditions, it was puzzling why the pilots failed to carry out the standard stall recovery procedure at the onset of the stall in this accident. The investigators wanted to know whether the engine-pylon separation contributed in any way to this failure.

It turned out that there was a causal path from C to H. I mentioned earlier that when the investigators examined the aircraft's electrical systems recovered from the wreckage, they discovered that the separation of the engine-pylon assembly damaged certain electrical wire bundles, which caused some (but not all) flight instruments in the cockpit to lose power. The flight instruments that lost power also included the stall warning systems. ([8], p.44) Let us say that event C was causally sufficient for **the stall warning systems in the cockpit to be disabled** (call this event **X**, abbreviated description "stall warning systems disabled").

Event X, together with the condition S (that the pilots did not realize that the stall speed of the left wing had increased), ensured that **the pilots did not recognize the onset of the stall** (condition **Y**, abbreviated description "ignorance of stall onset"), which was causally sufficient for H (the pilots did not follow standard stall recovery procedure). We can represent the above results as:

$$C \Rightarrow X,\ \{S, X\} \Rightarrow Y,\ Y \Rightarrow H.\ ^{17}$$

These results contain a causal path from C (engine separation) to H (no stall recovery), which was what the investigators were looking for. That marked the end of the investigators' process tracing between event C (engine separation) and event E (crash).

In sum, the investigators of the American Airlines Flight 191 accident identified three causal paths from event C (engine separation) to event G (stall) and one causal path from event C to event H (no stall recovery). Since both G and H were causally relevant to E (crash), it follows that the investigators identified a total of four causal paths from C to E. The discovery of these causal pathways helped the investigators gain a full picture about *how* event C contributed causally to event E.

Moreover, the consistent uses of sufficiency sets and INUS conditions in carrying out process tracing also gave the investigators ample opportunities for interventions: The removal of any INUS condition from a sufficiency set renders the remaining set *insufficient*; the disabling of any sufficiency relation along a causal path severs the entire path; and cutting off just *one* causal path between C and E may be sufficient to prevent E from happening even if C occurred.

In the American Airlines 191 investigation, the investigators identified multiple INUS conditions along the causal paths from C (engine separation) to E (crash) that could and should be changed in the future. These INUS conditions became the basis of a variety of safety recommendations. For instance, the lack of a mechanical slat locking device prompted

---

[17] Reminder: C (engine separation), S (ignorance of stall speed increase), X (stall warning systems disabled), Y (ignorance of stall onset), H (no stall recovery).

the NTSB to recommend a redesign of the slat control system to incorporate such a device. Further, The NTSB recommended a re-evaluation of American Airlines' engine-out emergency takeoff procedure, and a re-calculation of the speed schedules for the engine-out climb to ensure that they provide the maximum possible protection from the stall. Finally, the NTSB called for design changes that could enhance the redundancy of the stall warning systems. ([8], p.70-72) The implementation of these recommendations ensured that, even if a similar engine-pylon assembly separation happened again in the future, a stall and a subsequent crash would be much less likely to ensue.

Chapter 12

Conclusion to Part II

12.1 Recap

Let us recap the major findings of the preceding chapters.

In Chapter 8 ("Events, Features and Traces"), I distinguished between three main types of features of an event: Some properties of the entities involved in the event, the identities of the entities involved in the event, and the spatial and temporal characteristics of the event. I also proposed the following account of traces of an event:

A *trace* of a past event E is a property P of an entity X, such that the following three conditions are satisfied:

1. X exists at the time of the investigation with property P, and P is epistemically accessible to the investigators.
2. P provides epistemic access to some feature F of event E.
3. P provides epistemic access to feature F, either because P is identical with F, or because P is causally dependent on F and the systematic dependence relationship that connects F and P is known.

Traces allow the investigators to gain knowledge about features of at least some subevents[1] of the complex event to be reconstructed. This knowledge serves as the starting point of event reconstruction, and it can take the form of *premises* in reverse causal inference methods such as feature dependence and additional outcomes.

Chapter 9 ("Feature Dependence") discusses a type of reverse causal inference called "feature dependence" or "feature dependence arguments". This type of methods typically takes the features of some *known* subevent E (of the complex event to be reconstructed) as given, and infers the features of some other *hypothetical* subevent C, on the assumption that C was a cause of E.

A fundamental constituent of feature dependence arguments is a type of statements called *feature necessity*, and I proposed two formulations of these statements:

- Formulation 1 of feature necessity: If event E has a feature F_E, then, there must exist a cause C of E, such that C has a feature F_C.
- Formulation 2 of feature necessity: If event C is a cause of event E, then, given that E has a feature F_E, C must have a feature F_C.

I further explained the evidential basis of feature necessity statements using the idea of *event-based causal systems*. An event-based causal system is a type of causal system that instantiates whenever one type of event causes another type of event, and it supports systematic causal relationships among the feature types of these event types. Using this idea, we can depict the reasoning in support of Formulation 2 of feature necessity as follows:

- An actual causal relationship between C and $E \Rightarrow$ An instantiated event-based causal system $\Rightarrow$ A type-level causal relationship within the system that connect feature types of two event types $\Rightarrow$ A particular instance of the type-level causal relationship that connects feature F_C of event C and feature F_E of event E. [2]

---

[1] I use the term "subevent" whenever I want to emphasize the fact that the event in question is a part of a larger, more complex event to be reconstructed. A subevent is an event in itself, of course.

[2] Here "$\Rightarrow$" means an inference from one step of the reasoning to the next step.

Finally, I discussed a subtype of feature dependence argument called "descriptive enrichment". The basic idea of descriptive enrichment is that it combines multiple features of a subevent E to infer a detailed description of a hypothetical cause C of E. Here is an abstract representation of one version of descriptive enrichment:

- Premise 1: Event E has feature F_E^1, F_E^2, etc.
- Premise 2: If event E has a feature F_E^1, then, there must exist a cause of E that has a feature F_X^1.
- Conclusion 1: There exists an event that has feature F_X^1 and is a cause of E.
- Premise 3: If **this** event that has feature F_X^1 is a cause of E, then, given that E has another feature F_E^2, **this** event must have another feature F_X^2.
- Conclusion 2: There exists an event that has features F_X^1 and F_X^2, and it is a cause of E.
- Continue the process to learn more features about **this** event, until we can run additional causal arguments (such as additional outcomes) based on what we know about *this* event.

Chapter 10 ("Additional Outcomes") discusses a second type of reverse causal inference called "additional outcomes" or "additional outcomes arguments". This type of method evaluates the claim that a hypothetical subevent C caused a known subevent E, by determining whether C caused an *additional* outcome E^* distinct from E.

I made a distinction between the *negative* version and the *positive* version of additional outcomes arguments. The negative version of the argument denies that the hypothetical subevent C caused the known subevent E, and it has the following structure:

- Premise 1: If event C had occurred, there would have been some additional event E^* caused by C (or by some causal ancestor of C).

- Premise 2: Event E^* had not occurred.
- Conclusion 1: Event C had not occurred.
- Conclusion 2: It is not true that event C was a cause of the known event E.

In contrast, the positive version of the argument confirms that the hypothetical subevent C caused the known subevent E, and its structure can be represented as follows:

- Premise 1: Event E^*, an additional outcome different from event E, had occurred.
- Premise 2: E^* must (or likely) have been caused by the event C.
- Conclusion 1: Therefore, C must (or likely) have occurred.
- Candidate Premise 1: C was *capable* of causing E.
- Candidate Premise 2: (It is likely that) event E (the primary outcome) and event E^* (the additional outcome) had a common cause.
- Candidate Premise 3: For every other hypothetical event C^* that *has been considered* as a possible cause of E, there is evidence that either it had not occurred, or it was *incapable* of causing E.
- Conclusion 2: C must (or likely) have been a cause of E.

Additional outcomes can be used independently of feature dependence, but the two types of reverse causal inferences can also be combined together. For instance, we can use feature dependence arguments to infer the features of a hypothetical subevent C, and then evaluates whether C would cause certain additional outcomes given that it had such and such features. More concretely, consider the following (alternative) version of descriptive enrichment:

- Premise 1: Event E has feature F_E^1, F_E^2, etc.

- Premise 2: If an event of type X is a cause of event E, then, given that event E has a feature F^1_E, then, **this** event of type X has a feature F^1_X.
- Conclusion 1: If an event of type X is a cause of event E, then this event has feature F^1_X.
- Premise 3: If **this** event of type X is a cause of event E,, then, given that E has another feature F^2_E, **this** event must have another feature F^2_X.
- Conclusion 2: If an event of type X is a cause of event E, then this event has feature F^1_X and F^2_X.
- Continue the process to learn more features about **this** event of type X, *assuming* that it is a cause of E.

Note that the conclusion of this version of descriptive enrichment is a conditional statement: "If a subevent C of a type X was a cause of subevent E, then C had such and such features". We can evaluate the antecedent of this conditional statement using an additional outcomes argument, for instance (using a negative version of the argument):

- If an event C of type X with features F^1_X, F^2_X, etc had occurred, it would have caused some additional outcome E^*.
- Event E^* did not occur.
- Therefore, an event C of type X with features F^1_X, F^2_X, etc did not occur.
- If an event C of type X had caused event E, then C would have features F^1_X, F^2_X, etc. (This is the conclusion of the descriptive enrichment argument.)
- Therefore, it is not true that an event C of type X caused event E.

Chapter 11 ("Process Tracing") discusses a causal inference method called "process tracing" that combines elements of both forward and reverse causal inference. Process tracing establishes the causal connection between two subevents C and E of a complex event, by identifying causal pathways (consisting of causally connected mediating subevents) from C to E. It is commonly used when the investigators already know many of the subevents comprising the complex event to be reconstructed. It helps the investigators to "connect the dots" and obtain a fuller causal picture.

I argued that one away of carrying out process tracing makes systematic uses of the concept of INUS conditions. The basic idea is that whether we are tracing forward or tracing backward from a given subevent, we should always identify the entire set of INUS conditions that, taken together, is causally sufficient for the outcome. The main advantage of carrying out process tracing in this way is that given two target subevents C and E, we can identify not just one causal path from C to E, but all the causal paths from C to E in a particular case. This knowledge is useful because it gives us a better understanding of the underlying causal structure and reveals more opportunities for interventions.

12.2 A Few Comments About Success

Finally, I want to conclude with a few comments about the successes of reverse causal inference in engineering failure investigations (e.g., the NTSB's plane crash investigations).

One of my main reasons for using examples from engineering failure investigations to illustrate reverse causal inference is that these examples are very *successful*. By "successful", what I mean is that the instances of reverse causal inference in these examples are convincing, that the conclusions in each instance are supported by strong evidence, and that there is very little room for reasonable disagreement concerning these conclusions.

The causal inference methods I discussed (feature dependence, additional outcomes, and process tracing) are used not just in engineering failure investigations, but also in many other fields such as geology, paleontology, archaeology, and history. However, I believe that many examples of reverse causal inference in these fields are less successful than the examples I have discussed. The persistent scholarly disagreements on many causal conclusions (e.g., about the causes of the extinction of the dinosaurs) in these fields provide *prima facie* evidence in support of my view.

Suppose I am correct that on the whole, reverse causal inference in engineering failure investigations tend to be more successful that reverse causal inference in many other fields that are historical.[3] The question then becomes: Why is it the case? Is there anything special about engineering failure investigations that enable the reverse causal inferences to be more successful? A full discussion of these questions is outside of the scope of this dissertation. However, I believe that engineering failure investigations tend to possess certain epistemic characteristics that make it more likely for reverse causal inference to succeed. Below, I discuss two such epistemic characteristics.

First, engineering failure investigations tend to have access to traces that are better preserved and more informative. One reason for this is that the temporal distance between the event of interest and the investigation is relatively short, typically from a few hours to a few days. Because the investigation almost immediately followed the event of interest, many causal processes that could erode the traces did not have the time to operate. In contrast, historical sciences such as geology and paleontology study events that happened millions or billions of years ago. Subsequent causal processes have already destroyed many of the most informative traces from these events.

Another reason for the better quality traces in engineering failure investigations is that the investigators have some control over *which* traces are preserved and *how* they are preserved, and the control improves over time. For instance, one of the major obstacles encountered in the United 585 and USAir 427 investigations was that the FDR did not record enough flight parameters. After the investigations, the FAA significantly increased the

---

[3] Of course, this is an empirical claim, and I can be wrong about it. If I am wrong, I will simply narrow the scope of the statement, e.g., by saying that the engineering examples I have discussed are more successful than many examples of reverse causal inference in history and historical sciences. The question below then becomes whether any epistemic characteristics differentiate my examples from the many examples of reverse causal inference in history and historical sciences.

number of parameters required to be recorded by the FDR. So nowadays, the investigators have access to a lot of more traces of flight events. Similarly, the FDR and the CVR were *designed* to survive a crash, and the designs have only improved over time. In contrast, researchers in fields such as geology and paleontology have no control about which traces were preserved and how well they were preserved.

Second, engineering failure investigations tend to deal with a smaller set of causal systems that are better understood, compared to fields such as history. This epistemic characteristic contributes to the better evidential quality of feature dependence and process tracing used in engineering failure investigations.

For instance, in my discussion of the evidential basis of the feature necessity statement, I argued that in order to accept such a statement, we need to assume that a certain event-based causal system is at work, *and* no other event-based causal system that can modify the relevant systematic relationship among features is at work. Conversely, suppose we have reason to believe that multiple event-based causal systems are at work. In that case, we need to understand how they interact with each other, and the effect of their interactions on the systematic relationship among features in question.

Engineering failures typically involve a comparatively small set of causal systems, each of which is either relatively well understood or at least, can be identified, isolated and studied when necessary. In contrast, fields such as history study events that can involve a vast number of causal systems, most of which are not well understood. Consider events such as the Great Depression or the end of the Cold War. These events involve numerous causal systems, operating simultaneously at different levels of aggregation (individuals, groups, institutions, nations, etc), and neither the causal systems individually nor the interactions among them are well understood. Consequently, feature dependence arguments tend to be better supported in engineering failure investigations than in history.

Furthermore, in my discussion of process tracing, I mention that process tracing requires leveraging background knowledge about type-level causal sufficiency and causal necessity claims. Moreover, how convincing process tracing is depends on the strength of the type-level evidence for these causal claims, and the applicability of these type-level causal claims to the case in question. Since engineering failures typically involve a small set of relatively well understood causal systems, the causal generalizations relevant to the investigations are generally well supported by evidence (obtained primarily in laboratories). The investigators typically have an implicit understanding of the ceteris paribus clauses of these generalizations and whether these clauses are satisfied in a particular case.

In contrast, in fields such as history, robust causal generalizations are difficult to find. The sheer number of causal systems that could be operating concurrently also makes it very difficult to judge whether a given causal generalization applies to a particular case. Consequently, It is difficult to find examples of process tracing in history or social sciences that are as convincing as the American Airlines Flight 191 example I discussed earlier.

In sum, the above discussion, while by no means comprehensive, suggests that engineering failure investigations tend to possess certain epistemic characteristics conducive to successful reverse causal inference. A more in-depth comparison between engineering failure investigations and other types of event reconstruction research may identify more epistemic characteristics conducive to successful reverse causal inference, which in turn are crucial for a *normative* account of reverse causal inference.[4]

---

[4] My account of reverse causal inference in this dissertation is primarily a descriptive account. A normative account of reverse causal inference should identify general factors that determine how successful any given instance of reverse causal inference is.

Part III

Question Dynamics and Coherence

Chapter 13

Introduction to Part III

Part II of the dissertation focuses on how the investigators in an event reconstruction research discover causal relationships among subevents of a complex past event and support each causal conclusion with evidence. However, an individual causal conclusion is only a *very* small part of the output of an event reconstruction research. What an event reconstruction research ultimately produces is an entire *narrative* of the complex event, which details the full causal structure of the event. The main questions that I address in Part III of the dissertation are: How do investigators come up with an entire narrative of a complex past event? How do they support such a narrative with evidence? If we think of Part II as a *micro* study of event reconstruction research, then Part III is a *macro* study of event reconstruction research.

So, just like in Part II, I address both a question about discovery ("how do investigators come up with narratives of past events?") and a question about justification ("how do investigators support narratives of past events with evidence?") in Part III. Moreover, I believe that the two questions are closely connected and are not easily separated. My main thesis is that concerning narratives of past events, we can make substantial progress on both the question about discovery and the question about justification by examining the question-and-answer process in the investigation, which I call the *question dynamics*.

Consider first the question about discovery: “How do investigators come up with narratives of past events?” Narratives of past events produced by event reconstruction research tend to be highly complex, and investigators do not simply come up with these narratives out of nothing. Rather, the investigators answer one question at a time during the investigation. Eventually, they will have answered enough questions to be able to construct a narrative of the past event. Moreover, investigators do not ask questions randomly during the investigation; when they raise a question, they typically have good reasons to do so. Finally, there is *structure* to the question-and-answer process, in that the investigators tend to ask questions that build on the answers to previous questions that they have already resolved. These are *prima facie* reasons that examining the structure of question dynamics helps us understand how investigators come up with narratives of past events.

Next, consider the question about justification: “How do investigators support narratives of past events with evidence?” More precisely, what I am asking is: “What makes a narrative more, or less, well supported by evidence?” Here, an intuitive answer is that a narrative that is better supported by evidence is a narrative that is more *coherent*, i.e., a narrative in which all the pieces “fit together” very well. Of course, since we are talking about event reconstruction research rather than fiction, a coherent narrative cannot just mean an internally consistent story; at the very least, a coherent narrative needs to be consistent with all the available traces left from the past event. Beyond this, how do we make the notion of coherence more precise?

I claim that the coherence of narratives of past events can also be explicated in terms of the structure of question dynamics in event reconstruction research. To motivate this idea, consider an analogy with how we solve jigsaw puzzles. The notion of a coherent narrative is analogous to a completed jigsaw puzzle where “all the pieces fit together”. One way to think about what “all the pieces fit together” means is to think about how we solve jigsaw puzzles sequentially. To produce a jigsaw puzzle where “all pieces fit together”, we just need to add to the puzzle one piece at a time. The earlier pieces that we put in provide *constraints* on how we add later pieces. Moreover, if we chose the earlier pieces correctly, these constraints can amount to *clues* about how the later pieces should fit. If we can continually add pieces to the puzzle given the constraints of the previous pieces, we will eventually complete the puzzle where "everything fits together".

Similarly, to create a coherent narrative, the investigators need to resolve one question at a time. The answers to questions they resolved earlier provide both clues about, and constraints on, what questions they should ask next, and possible answers to these questions. If the investigators can continue to raise and resolve new questions by building on the answers to previously resolved questions, they will eventually be able to reach a point where they have put "all the pieces together". This analogy provides a *prima facie* reason that examining the structure of question dynamics helps us understand what it means for a narrative of a past event to be coherent.

The main task of Part III of this dissertation is to explicate the structure of question dynamics in event reconstruction research. Upon inspection, we can see that question dynamics has three major aspects. First, questions have to be resolved in question dynamics. Otherwise, nothing is achieved by raising questions. Second, different questions have different degrees of significance for the event reconstruction research, and for the narrative that it produces. Finally, question dynamics is a dynamic process, because questions continue to arise even while some are resolved. These three main aspects of question dynamics correspond to the three chapters in this part of the dissertation.

Part III is organized as follows. First, in chapter 14 ("Question Resolution"), I examine the resolution of questions. I begin with an abstract account of what it means to resolve a question. To situate this abstract account in event reconstruction research, I focus specifically on questions that arise in event reconstruction research, which I call "narrative questions". I discuss the main forms of narrative questions, what constitute candidate answers to these forms of questions, and the forms of evidence for or against these candidate answers.

Second, in chapter 15 ("Question Significance"), I examine the significance of questions. I begin with a review of Sylvain Bromberger's distinction among four types of values of questions. Based on this framework, I make a similar distinction among four types of the significance of questions and examine each type of question significance in turn: The Peircean significance of a question is its intellectual significance; the Jamesian significance of a question is its practical significance; the Machian significance of question consists of its potential for answering other questions; and the Collingwood significance of a question measures the impact of its resolution on the entire narrative.

Third, in chapter 16 ("Question Dynamics and Coherence"), I begin by examining what it means for questions to arise, and what it means for something to give rise to questions. After rejecting the semantic account of what gives rise to questions common in the existing literature, I propose my account of what gives rise to questions, which uses my account of question significance. Next, using all the conceptual resources I have introduced so far, I describe the full structure of question dynamics in event reconstruction research. Last, I propose an account of coherence of narratives based on structural parameters in the questions dynamics of event reconstruction research.

Finally, I conclude part III of the dissertation with a recap of the major findings. I also make a few comments about the extent to which my account of question dynamics and coherence applies to event reconstruction research in general, given that all my examples are drawn from a particular type of event reconstruction research, namely engineering failure investigations.

Chapter 14

Question Resolution

14.1 Question Resolution: An Abstract Account

An investigation that reconstructs a complex past event is driven by the arising and the resolution of questions. At each point within the investigation, some questions are resolved, at least provisionally, whereas new questions may arise simultaneously. In this chapter, I examine how questions are resolved in event reconstruction research, using examples from engineering failure investigations. The topic of how questions arise will be discussed in a later chapter.

I begin with a simple and abstract account of what it *means* to resolve a question that arose in an investigation. At a given time t within the investigative process, any question Q that has risen by time t has a set $\mathcal{A}_Q$ of *candidate* answers to it. By "candidate answers", I mean possible answers that (1) are epistemically accessible[1] to the investigators at time t and (2) cannot be ruled out based on the evidence available to the investigators at time t. They are answers that, for all that the investigators know at time t, can still turn out to be the correct answer to Q. To simplify the presentation, I assume that each candidate answer

---

[1] To say that a possible answer A (to a question Q) is epistemically accessible to the investigators at time t, is to say that based on their conceptual resources and background knowledge at time t, the investigators are *in a position* to conceive and to consider that possible answer A.

A in the set $\mathcal{A}_Q$ is a *full* answer[2] to the question Q, and that different candidate answers are mutually incompatible—that is, for any two different candidate answers in $\mathcal{A}_Q$, at least one of them must be an incorrect answer to Q. At a given time t within the investigation, the set $\mathcal{A}_Q$ can be empty, a singleton, or a set with multiple members.

Given this setup, we can define *open* and *resolved* questions as follows:

- A question Q is *open* at time t, just in case the corresponding set $\mathcal{A}_Q$ of candidate answers is either an empty set or a set with multiple elements at time t.
- A question Q is *resolved* at time t, just in case the corresponding set $\mathcal{A}_Q$ of candidate answers is a singleton at time t.

The key feature of this definition is that the resolution of a question is defined relative to time. It allows the possibility that a question Q is resolved at time t_1, and yet becomes open again at a later time t_2. In other words, what is being defined here is a notion of *provisional* resolution of questions. Nevertheless, defining a notion of permanent resolution of questions is also straightforward: A question Q is permanently resolved, just in case it is resolved at time t and remains resolved at all times after t.

How could it be that a question Q is resolved at time t_1, and yet becomes open again at a later time t_2? There are multiple possibilities. First, the set $\mathcal{A}_Q$ of candidate answers to Q may increase in size over time, because certain possible answers that were previously inconceivable or inaccessible to the investigators may become available at time t. For instance, suppose that Q was resolved at time t_1, and the set $\mathcal{A}_Q$ of candidate answers was a singleton $\{A_1\}$ at t_1. However, at a later time t_2, the investigators discover another possible answer A_2 to question Q, which did not occur to them before and cannot be ruled out by evidence available at t_2. The set $\mathcal{A}_Q$ of candidate answers to Q at t_2, then, is $\{A_1, A_2\}$, which is no longer a singleton. Hence the Q becomes open again at t_2.

---

[2] The reason I stipulate that candidate answers are full answers is that partial answers to a question do not have to be mutually incompatible.

Second, the amount of available evidence for or against each possible answer also changes over time. Possible answers that were ruled out before may no longer be ruled out given new evidence. For instance, suppose that Q was open at time t_1, and the set $\mathcal{A}_Q$ of candidate answers to Q was $\{A_0, A_1\}$ at t_0. At a later time t_1 ($t_1 > t_0$), the investigators found enough evidence[3] in support of A_1 and enough evidence against A_0, that they decided that A_0 could be ruled out. So $\mathcal{A}_Q$ became a singleton $\{A_1\}$ at t_1, and Q was resolved at time t_1. However, at a later time t_2 ($t_2 > t_1$), the investigators discover new evidence against A_1 and in support of A_0, which means that they can no longer rule out A_0. Hence $\mathcal{A}_Q$ is $\{A_0, A_1\}$ at t_2, and Q becomes open again at time t_2.

This completes my account of what it means to resolve questions that arise in an investigation. This account is sufficiently abstract to apply to many different types of inquiries. As such, it is not specific enough to be informative about the resolution of questions in the reconstruction of past events. For instance, what *types* of questions typically arise when investigators reconstruct complex events of the past? What constitutes candidate answers to these types of questions, and what forms of evidence do the investigators use to evaluate the candidate answers? In the rest of this chapter, I address these philosophical questions by examining examples from engineering failure investigations.

14.2 Narrative Questions and Their Types

A distinctive feature of event reconstruction research is that the answers to the questions that arise within the investigative process can be used to construct and support a *narrative* of the past event. Let us define *narrative questions* as those questions whose answers constitute part of the narrative of a past event. Many—if not all—questions that arise in

---

[3] What counts as "enough evidence" is a matter of judgment. In practice, the investigators examine the balance of evidence for and against each candidate answer and compare the overall evidential strength of different candidate answers. If the overall evidential strength of two candidate answers differs sufficiently (again a matter of judgment), the investigators can be justified to rule out the one with weaker support.

an event reconstruction research are narrative questions.[4] In this section, I examine the common types of narrative questions in event reconstruction research.

What types of questions are narrative questions? In the rest of this dissertation, I address this philosophical question by examining the five cases of aviation failure investigations conducted by the National Transportation Safety Board. Unfortunately, the NTSB accident reports do not record the questions that the investigators raised during the investigation. Instead, the reports only document the answers to the narrative questions and the evidence in support of these answers. This means I will have to reconstruct the narrative questions based on their answers, which I will call *narrative answers*.

A natural place to look for narrative answers obtained in an NTSB investigation is the 'Conclusion' section of the NTSB reports. A survey of this section of the reports showed three main types of narrative answers based on their content.

The first type of narrative answers are statements about the *absence* of various kinds of causal relationships such as causal necessity, causal sufficiency, or feature dependence, etc. For instance:

- "Weather was not a factor in this accident." ([11], p.100)
- "Galling found on the input shaft and bearing from the standby rudder actuator power control unit could not cause sufficient rudder deflection to render the airplane uncontrollable." ([12], p.102)
- "Although USAir flight 427 encountered turbulence from Delta flight 1083's wake vortices, the wake vortex encounter alone would not have caused the continued heading change that occurred after 1903:00." ([10], p.292)

---

[4] Sometimes, the investigators may raise questions that are not directly about the past event being reconstructed, and the answers to these questions are not included in the narrative of the event. For instance, in an NTSB investigation, sometimes the investigators may discover some vulnerability or risk factors, and they may raise questions about how dangerous these factors are and what should be done about them, even though these vulnerabilities and risk factors were not involved in the accident.

The second type of narrative answers are statements about the *presence* of various kinds of causal relationships. For instance:

- "Separation of the titanium alloy stage 1 fan rotor disk was the result of a fatigue crack that initiated from a type 1 hard alpha metallurgical defect on the surface of the disk bore." ([11], p.100)
- "A short circuit producing excess voltage that was transferred to the center wing tank (CWT) fuel quantity indication system wiring is the most likely source of ignition energy for the TWA flight 800 CWT explosion." ([13], p.307)
- "A fuel/air explosion in the center wing fuel tank of TWA flight 800 would have been capable of generating sufficient internal pressure to break apart the tank." ([13], p.306)

Finally, the third type of narrative answers are statements about the presence or absence of certain hypothetical subevents, background conditions, or features. For instance:

- "At the time of the accident, there were light winds and scattered clouds in the area, but there were no significant meteorological conditions that might have disrupted the flight." ([13], p.306)
- "A detectable fatigue crack about 0.5 inch long at the surface of the stage 1 fan disk bore of the No. 2 engine existed at the time of the most recent United Airlines inspection in April 1988 but was not detected before the accident." ([11], p.102)
- "About 1903:00, USAir flight 427's rudder deflected rapidly to the left and reached its left aerodynamic blowdown limit shortly thereafter." ([10], p.292)

In sum, all three types of narrative answers are statements about the **causal structure** of the complex event to be reconstructed. By "causal structure", I mean (1) the subevents that constitute the complex event and features of these subevents, (2) the background conditions that made some causal contributions at some point during the complex event, and (3) the presence or the absence of various causal relationships among the subevents, the background conditions and the final outcomes of the event.

Since the narrative answers typically describe some aspects of the causal structure of the complex event, narrative questions must be *queries* about that causal structure. Moreover, it is relatively straightforward to infer the common *forms* of narrative questions: They include why-questions, how-questions, and wh-questions[5].

First, when the investigator identifies an abnormal[6] trace or infers an abnormal subevent, they often raised *why* questions: Why this abnormal trace was there, or why the abnormal event had occurred. For instance:

- At the beginning of the Flight 191 investigation, both eye-witnesses of the accident and the wreckage distribution suggested that the No.1 (left) engine had separated from the left wing during takeoff. The investigators raised a question that could be paraphrased as: “Why did the No.1 engine fall off?” ([38], p.9)
- During the United 232 investigation, the investigators established that (1) there was a fatigue crack about 1/2 inch long along the bore surface of the stage 1 fan disk during the last Fluorescent Penetrant Inspection (FPI) of the disk before the accident, but (2) the inspectors who carried out the inspection failed to detect the crack. ([11], p.87) Given that such a crack should have a high probability of detection by FPI, the investigators raised the question: “Why didn’t the inspectors detect the fatigue crack during those inspections when the crack should have been detectable?”
- Also during the United 232 investigation, the investigators found a discolored area on the fracture surface of the fatigue crack. ([11], p.85) They raised the question: “Why

---

[5] Wh-questions include what-questions, which-questions, where-questions, and when-questions.

[6] By “abnormal” I mean “contrary to expectations”.

was there such a discolored area on the fatigue fracture surface?"

Second, closely related to the why-questions are the *how*-questions. For instance:

- In the TWA 800 investigation, the wreckage distribution suggested that the wing center section of the fuselage failed first. One subgroup of investigators, the sequencing group, addressed the question "*How* did the wing center section fail?". This group's main task was to determine the sequences of events leading from the initial failure to the separation of the wing center section from the aircraft. ([14])
- In the USAir 427 investigation, The FDR data indicated a highly abnormal subevent in the failure sequence. Namely, the accident aircraft yawed and rolled abruptly to the left and continued to do so until ground impact. Based on flight simulations, the NTSB concluded that the only thing that could make the 737 yaw and roll to the left at the recorded rate was a rudder hardover event—the rudder deflected to the extreme position (called the blowdown limit), full left in this case. ([10], p.59) A main possibility pursued by the NTSB was that a jam inside the main rudder power control unit (PCU) caused the rudder hardover event. However, common forms of jam (e.g., by particles or metal chips) typically leave behind telltale witness marks on the jammed component. Yet, examinations of the inside of the main rudder PCU found no witness marks. ([10], p.73-76) For a few years, the NTSB investigators grappled with the question: "Could the main rudder PCU of the accident aircraft jam without leaving any witness marks behind? If so, *how*?"

Third, causal questions can also be phrased as *what* or *which*-questions. On some occasions, the investigators knew that in order for an abnormal subevent or trace X to occur or exist, *something* must have played a contributing causal role R; but they did not know the identity of the object that had played the causal role R. This is a situation where what-questions can be appropriate. For instance:

- At one point in the TWA 800 investigation, NTSB was convinced that a fuel/air explosion in the center wing tank (CWT) caused the inflight breakup. Even assuming that the fuel/air mixture inside the tank was flammable, for it to be ignited, there must be an ignition source (i.e., something must have provided the fuel/air mixture sufficient ignition energy to ignite it). This gave rise to the following question: "What ignited the inflammable fuel/air vapor in the CWT?" ([13], p.271)

Moreover, why-questions are sometimes paraphrased as "what caused" questions, as in "What caused event E"? For instance, the question "why did the fuel/air mixture inside the center wing tank explode?" is sometimes paraphrased as "what caused the explosion of the fuel/air mixture inside the center wing tank?".

What-questions, in turn, can become which-questions, if only a few options could realistically have played the causal role R or caused the outcome E. In such a situation, the investigators could ask *which* of the candidates played the causal role. For instance:

- In the United 232 investigation, the investigators established through chemical analysis that the discoloration of the fatigue fracture surface was due to a FPI (Fluorescent Penetrant Inspection). ([11], p.85) Since the accident fan disk only went through a total of six FPI inspections in its lifetime, the investigators could ask: "Which of the six FPI caused the discoloration?"

Finally, the investigators would raise "when" and "where"-questions, if they are interested in certain temporal and spatial features of subevents or background conditions. For instance:

- In the United 232 investigation, the investigators found that the fatigue crack originated from a hard alpha area, and a cavity at the center of the hard alpha area. ([11], p.45) They examined the question about *when* the hard alpha area, and *when* the cavity at the center of the hard alpha area, were formed.

- In the USAir 427 investigation, the investigators identified a few mysterious thumping sounds in the CVR recording, right at the initiation of the failure sequence. To determine the source of thumping sounds, the investigators first addressed the question about *where* the source of the thumps was using spectral analysis. ([18], p.175)

In sum, narrative questions are questions about the causal structure of the complex event to be reconstructed, and they can take the form of why-questions, how-questions, and other wh-questions. In the next section, I examine the *forms* of narrative answers; i.e., what constitutes answers to the narrative questions such as why-questions and how-questions.

14.3 Narrative Answers

We have seen that the *content* of narrative answers—i.e., answers used to construct a narrative of the past event—consists of details of the causal structure of that event. But what about the *forms* of narrative answers? Given that narrative questions can take the forms of why-questions, how-questions, and other wh-questions, what count as candidate answers to these questions?

For wh-questions such as when, where and which-questions, it is relatively straightforward to determine what counts as candidate answers to these questions. For instance, for a question "where did subevent E occur?", a candidate answer to this question is simply a possible *location* that (1) the investigators could come up with and (2) could turn out to be the true spatial location of subevent E given what the investigators know. Consequently, in this section, I focus on why-questions and how-questions, and what constitutes candidate answers to these questions in the context of event reconstruction research.

What counts as a candidate answer to a why-question that arises in an event reconstruction research? My proposal is derived from van Fraassen's contrastive theory of why-questions. In a nutshell, I propose that a candidate answer to a why-question describes a causal factor that (purportedly) makes a specific difference.

According to van Fraassen, the correct underlying structure of a why-question is contrastive. Formally, van Fraassen analyzes a why-question Q as a triple $\langle P, X, R\rangle$. First, P, the *topic* of the question, is a proposition expressing the fact whose explanation we are seeking. Second, X, the *contrast class*, is a set of propositions including the topic, and it also includes the contrasts against which the why-question is asked. Finally, R, the *explanatory relevance* relation, is a relation that the answer should bear to P and X; it plays a selective role by determining what counts as an explanatory factor.([37], p.141-2) Presumably, both the contrast class and the explanatory relevance relation are supplied by context.

Given this contrastive theory of why-questions, van Fraassen proposes that a direct answer to a why-question $Q = \langle P, X, R\rangle$ takes the following form:

> P in contrast to (the rest of) X because A.

where A is a proposition that bears the relevance relation R to the topic P and the contrast class X. Moreover, van Fraassen interprets the word "because" as being truth-functionally equivalent to "and".([37], p.144) That is, if we take the truth value of P and those of the elements of X for granted, we can simplify the answer to just "A", where A is a proposition that bears relation R to the pair $\langle P, X\rangle$.

The abstractness of van Fraassen's account reflects the fact that it is a general account of why-questions and their answers. In contrast, my concern is to understand why questions and their answers *in the context of event reconstruction research.* In this context, why-questions possess a few additional characteristics.

First, why-questions in event reconstruction research typically arise from *anomalies.* For instance, in the USAir 427 investigation, the FDR data indicated that the accident sequence began at about 19:02:58, when the accident aircraft began to yaw and roll at an increasing rate to the left. By 19:03:01, the airplane's heading was moving left at least 5 degrees per second until the aircraft stalled at 19:03:08. ([10], p.4-6) This highly abnormal behavior gave rise to the following why-question Q: "Why did USAir 427 yaw and roll so drastically during this period?" In van Fraassen's terminology, the topic P of the why question Q is the proposition "USAir 427 yawed and rolled at a high rate from 19:02:58 to 19:03:08", which describes an abnormal event.

Second, the contrast class X typically consists of propositions that describe *normal* events or states of affairs, in contrast with the anomaly described by P; and *which* normal event or state of affair is included in X depends on the context. Before the initiation of its yaw and roll, USAir 427 was coming out of a left turn towards a wing level attitude at 19:02:52, as it approached the heading (100 degrees), airspeed (190 knots) and altitude (600 feet msl) assigned by the Air Traffic Control (ATC). One possible contrast P_0 is that USAir 427 did not initiate yaw and roll at all, but rather continued its wing level flight at ATC-assigned parameters. Another possible contrast P_1 is that USAir 427 initiated its yaw and roll movement at 19:02:58, but quickly broke out of the yaw and roll and regained wing level flight. Both of these contrasts can be regarded as normal or expected in contrast with P. If the investigators were primarily interested in the contrast class $\{P, P_0\}$, their why-question Q could be paraphrased as "why did the yaw and roll movement *initiate*?". If the investigators were primarily interested in the contrast class $\{P, P_1\}$, their why-question Q could be paraphrased as "why did the yaw and roll movement *persist* after the initiation?".[7]

Third, the explanatory relevance relation R is a *difference-making* relation: The why-question requests a causal factor that makes a specific difference, namely that it resulted in P rather than its normal alternatives being true. The explanatory relevance relation R helps select a difference-maker by determining which conditions are held fixed. For instance, suppose the contrast class is $\{P, P_1\}$ in a given context, and the why-question Q can be paraphrased as "why did the yaw and roll movement *persist* after the initiation?". Different causal factors might have contributed to the persistence of the yaw and roll, and the explanatory relevance relation can help select some of these factors as relevant by

---

[7] As it turned out, an encounter with wake turbulence caused the initiation of the yaw and roll, whereas a rudder hardover caused the persistence of the yaw and roll. Hence different contrast classes did result in different answers to the why-question.

holding the others fixed. For instance, perhaps the investigators were primarily interested in *mechanical* malfunctions that contributed to the persistence of yaw and roll. Alternatively, perhaps the investigators took the mechanical malfunctions for granted, and were instead looking for human factors that made a difference *given* the mechanical malfunctions.

Based on these additional characteristics, I propose the following account of answers to why-questions in event reconstruction research. Let a why-question Q be a triple $\langle P, X, R\rangle$, where P describes an anomaly, X is the contrast class that includes P and its normal alternatives, and R is a difference-making relation. To simplify the presentation, let $X = \{P, P_0\}$, where P_0 is a normal alternative to P salient in a given context. A *correct* answer to question Q is a true statement A that bears the difference-making relation R to P and P_0, in the following sense:

- First, if A had been true (which it is), P instead of P_0 would have been true.
- Second, if A had been false, P_0 instead of P would have been true.

The precise truth conditions of these two counterfactual conditionals depend on the difference-making relation R, which determines which background conditions should be held fixed when evaluating these counterfactuals. Finally, a *candidate* answer to question Q at a given time t is a statement that, given what the investigators know at that time, could turn out to be the correct answer to Q.

In short, candidate answers to why questions that arise in event reconstruction research are statements that describe purported difference makers. What counts as a difference-maker, and which difference-maker is requested, may vary from context to context. Note that a statement describing a difference-maker can be a full answer to a why question, even though the difference-maker by itself is insufficient to cause the outcome.

Next, what counts as a candidate answer to a how-question that arises in an event reconstruction research? In my view, how-questions are generally more demanding than why-questions in event reconstruction research. To answer a why-question when reconstructing past events, we simply need to cite an appropriate difference-maker. To answer a how-question, in contrast, we typically give an account of the causal processes that led to an outcome, which tends to be a more difficult task and does more work for filling in the details for a narrative of the past event.

To motivate this difference between why-questions and how-questions, consider an example from linguistics that does not deal with reconstructing a past event.[8] If I ask a linguist: "Why could a child learn a language despite the poverty of the input?", the answer could simply be: "a universal grammar". A more detailed answer will specify detailed structures of universal grammar, which linguists have been working on for decades and have made significant progress.

On the other hand, if I ask: "How did the child learn a language despite the poverty of the input?", then "universal grammar" will not be sufficient at all as an answer. To answer the how-question, we need to detail a causal and developmental process, and explain how a universal grammar plus the empirical inputs helps generate a language. This is a more difficult project in which the linguists have had much less success so far.

Returning to how-questions in event reconstruction research, the basic ideas of my proposal about their answers are the following:[9]

- The question "How did X occur?" asks for the causal processes that actually produced X.
- The question "How could X (possibly) occur?" asks for any possible causal processes that could produce X.

---

[8] This example is due to George Smith, although he is not responsible for my presentation of the example or any possible error in it.

[9] These two forms are not the only possible forms that how-questions can take in event reconstruction research. Other possible forms include: "How did X occur given (or despite of) Y?", or "How did Y contribute to X?", etc. Nevertheless, I believe that it is relatively straightforward to extend my account to most of these alternative forms of how-questions.

By “causal processes that produce X”, I mean (1) a collection $\mathcal{C}$ of events and factors that, taken together, are *causally sufficient* for X to occur; (2) the sequences of (intermediate) events leading from the events and conditions in $\mathcal{C}$ to the outcome X. To answer a “how did” question, we need to specify causal processes that actually existed. To answer a “how could” question, we only need to specify causal processes that are possible in some relevant sense.[10] Note my emphasis on causal sufficiency: In my view, unlike why-questions, how-questions ask for *sufficient conditions* and the processes through which these conditions bring about an outcome.

To illustrate my account of how-questions, consider the USAir 427 example again. I mentioned earlier that based on flight simulations, the NTSB concluded that the only thing that could make USAir flight 427 yaw and roll to the left at the recorded rate was a rudder hardover event—the rudder deflected to the extreme position (called the blowdown limit), full left in this case. A main possibility pursued by the NTSB was that a jam inside the main rudder power control unit (PCU) caused the rudder hardover event. However, common forms of jam (e.g., by particles or metal chips) typically leave behind telltale witness marks on the jammed component. Yet, examinations of the inside of the main rudder PCU found no such witness marks. ([10], p.69-76) This gave rise to the question: “How could the main rudder PCU of the accident aircraft jam without leaving any witness marks behind?”

For over a year after the accident, the NTSB investigators failed to identify any plausible jamming mechanism inside the main rudder PCU that could produce enough binding force, while leaving behind no witness marks. Eventually, during a meeting by an independent technical advisory panel, an outside panel member suggested that thermal shock could be the mechanism that the investigators were looking for. A thermal shock occurs when overheated hydraulic fluid (typically heated by malfunctioning hydraulic pumps) flows into the cold body of the dual concentric servo valve inside the main rudder PCU, causing thermal expansion of the two moving slides within the servo valve. The thermal expansion

---

[10] The sense of possibility could be a physical possibility for a particular type of physical systems, etc.

of the slides could potentially result in a jam.[11] ([18], p.188)

To determine whether thermal shock could jam the dual concentric servo valve without leaving any witness marks behind, the NTSB investigators carried out a series of thermal shock tests on the main rudder PCU from USAir 427 and a new main rudder PCU. They discovered that unlike the new PCU, the main rudder PCU from USAir 427 could jam due to thermal expansion caused by thermal shock. More precisely, the secondary (outer) slide inside the dual concentric servo valve would jam to the servo valve housing due to thermal shock. ([10], p.78-79) Moreover, post-test examinations showed that the jam would not leave behind any witness marks on the servo valve housing or the slides. ([10], p.245-246)

In addition to thermal shock, the investigators further identified other necessary conditions for the jam: First, the dual concentric servo valve in the main rudder PCU of USAir 427 had tighter clearances than normal, which made it more prone to jamming. The new rudder PCU, in contrast, did not jam during the tests. Second, the jamming only occurred under an extreme temperature differential condition—a temperature difference of over 200 degrees Fahrenheit between the heated hydraulic fluid and the cold PCU inlet, which rarely obtains in a normal flight. ([10], p.78-79) The investigators established that if there had been a thermal shock satisfying the extreme temperature differential condition, then, given the relatively small clearance of the dual concentric servo valve, the secondary slide would have jammed to the servo valve housing.

In short, after the thermal shock tests, the NTSB investigators were able to identify a set of events and conditions that *could* be sufficient for a jam of the main rudder PCU, and the causal process leading from these events and conditions to the jam is relatively straightforward to describe using physics. However, the investigators did not have direct evidence for the actual occurrence of all the events and conditions in the set. In other words, the investigators were able to answer the "how could" question, but were not yet able to answer the "how did" question at this point in the investigation.

---

[11] For details about the dual concentric servo valve, please refer to my case summary on USAir flight 427.

To summarize: In event reconstruction research, candidate answers to why-questions are typically statements that describe *difference makers*, whereas candidate answers to how-questions are typically statements that describe *causal processes* that are sufficient to bring out an outcome.[12] In the next section, I discuss the common forms of evidence that the investigators use to evaluate these candidate answers.

14.4 Two Types of Evidential Questions

As long as a narrative question Q still arises in an investigation, it is associated with a set $\mathcal{A}_Q$ of candidate answers. Even though each candidate answer in $\mathcal{A}_Q$ has at least some epistemic possibility[13] of being the correct answer to Q, some candidate answers in $\mathcal{A}_Q$ may be better supported by evidence than others. To resolve the question Q at a given time t—at least provisionally—is to possess a body of evidence at time t that supports one of the candidate answers in $\mathcal{A}_Q$ and also rules out all the other competing candidate answers in $\mathcal{A}_Q$. The resolution of a question can be more or less provisional depending on the strength of evidence: The stronger the evidence in support one of the candidate answer is, and the more definitive it rules out existing and possible alternatives to that candidate answer, the less provisional a resolution is.

How do investigators obtain evidence for or against candidate answers to a narrative question? I propose that the investigators obtain evidence by raising and answering *further* questions, which I call *evidential questions*.[14] Let Q^N be a narrative question. For *each* candidate answer A^N to the question Q^N, the investigators can raise evidential questions about A^N. Moreover, each evidential question about A^N serves as a *test* of A^N, and the

---

[12] I say "typically" because there are exceptions: For instance, in the chapter on feature dependence, I mentioned that the answers to *some* how-questions can be descriptions of feature dependence—i.e., how certain features of event X make a difference to certain features of event Y.

[13] To say that a candidate answer A has an epistemic possibility of being the correct answer to question Q at time t is to say that, given what the investigators know at time t, A could still turn out to be the correct answer to Q.

[14] As we shall see, narrative questions and evidential questions are not mutually exclusive categories; the answers to some evidential questions are incorporated into the narrative of the past event.

structures of the evidential questions are informative about the *forms* of evidence for or against A^N.

Next, I distinguish between two types of evidential questions, which I call *evaluative* questions and *conditional* questions, respectively. Both of these two types of evidential questions are tests of a given candidate narrative answer. The major difference between evaluative questions and conditional questions is this: Evaluative questions about a candidate answer do not *presuppose* that candidate answer, so they can be meaningfully posed even if the candidate answer is known to be false. In contrast, conditional questions about a candidate answer presuppose that candidate answer, so they can no longer be meaningfully posed if the candidate answer is known to be false. As we shall see, because of this difference, the two types of evidential questions constitute different *forms* of tests of candidate answers.

Let me illustrate the distinction between evaluative questions and conditional questions using an example from the TWA 800 investigation. The sequencing study determined that the inflight breakup of TWA flight 800 was initiated by an overpressure event inside the center wing fuel tank. This gave rise to a narrative question Q_0: "Why did an overpressure event occur within the center wing tank"? One of the candidate answers to Q_0 that the investigators came up with, which I will call A_1, is that **some ignition source internal to the aircraft had ignited a flammable fuel/air mixture inside the center wing tank and caused an explosion**. Initially, the candidate answer A_1 was entirely hypothetical, and the investigators had little direct evidence for it. To collect more evidence on A_1, the investigators raised three major evidential questions about it: ([13], p.138)

1. Question Q_1: Was there a flammable mixture of fuel vapor and air inside the center wing tank of TWA Flight 800 at the time of the accident?
2. Question Q_2: Could the ignition and combustion of this flammable mixture of fuel vapor and air generate sufficient pressure to break apart the center wing tank?

3. Question Q_3: What was the ignition source inside the aircraft that ignited the flammable fuel/air mixture inside the center wing tank and caused an explosion?

Among the three questions, Q_1 and Q_2 are evaluative questions about candidate answer A_1, whereas Q_3 is a conditional question about candidate answer A_1. This is because neither Q_1 nor Q_2 presuppose A_1 (although Q_2 does presuppose a positive answer to Q_1): For instance, Q_1 does not presuppose that there was a flammable mixture of fuel/air vapor inside the center wing tank; rather, its point is to *evaluate* whether there was such a flammable mixture. Similarly, even though Q_2 does presuppose that there was a flammable mixture of fuel/air vapor inside the tank, it does not presuppose that the ignition of the fuel/air vapor would cause enough pressure to cause an explosion of the fuel tank.[15] Therefore, neither Q_1 nor Q_2 presupposes candidate answer A_1 in its entirety. Even if there was no flammable fuel/air mixture at the time of the accident, Q_1 is still a meaningful question—its answer is “no”. If there was a flammable fuel/air mixture, but its ignition could not cause an explosion, Q_2 is still a meaningful question, again with “no” as an answer.

In contrast, question Q_3 does presuppose candidate answer A_1, which states that *some* internal ignition source ignited the flammable fuel/air mixture inside the center wing tank and caused an explosion; Q_3 simply asks *what* that ignition source was. If A_1 is true, then Q_3 should have a correct answer, even if finding its correct answer can be challenging. If A_1 is false, however, Q_3 would not be meaningfully posable at all, and it certainly would not have a correct answer.

I would like to make four additional comments about this example. **First**, all the three evidential questions Q_1, Q_2 and Q_3 are empirical questions, in that they should not be answered in an *ad hoc* manner. Instead, investigators need to learn new empirical information to answer them. For instance, to answer Q_1 (whether there was a flammable mixture of

---

[15] The NTSB investigators considered the possibility that the ignition of fuel/air vapor resulted in a state of slow combustion, which would not generate enough pressure to break apart the fuel tank and cause an explosion.

fuel/air vapor in the center wing tank at the time of the accident), the investigators had to determine whether all the necessary conditions for a flammable fuel/air mixture were present at the time of the accident. For instance, one necessary condition is that the ratio between fuel molecules and air molecules in the vapor space[16] of the center wing tank must be within a certain range. Another necessary condition is that the temperature inside the center wing tank must be sufficiently high. To determine whether these necessary conditions were present, the investigators raised the following evidential questions:[17]

- Question Q_4: What was the ratio of fuel molecules versus air molecules in the center wing tank's vapor space at the time of the accident?
- Question Q_5: What was the temperature and thermal environment inside the center wing tank at the time of the accident?

These further questions, in turn, were investigated through post-accident flight tests. The investigators carried out tests flight that simulated the conditions of TWA 800 as close as they could. They measured the temperature of the center wing tanks during the test flight and analyzed the chemical compositions of the fuel/air vapor sampled from the fuel tanks.

Second, for evaluative questions about the narrative answer[18] A_1, their candidate answers can be divided into two groups: One group of candidate answers provide evidence *for* A_1, the other group of candidate answers provide evidence *against* A_1. Questions Q_1 and Q_2 are extreme cases for this: Both questions are binary questions with two possible answers only, "yes" and "no"; both ask whether a necessary condition for A_1 obtains or not. In both cases, the "yes" answer provides evidence for A_1 by making A_1 more likely, although it does not guarantee that A_1 is true. In contrast, the "no" answer provides evidence against A_1 in a stronger sense, because it guarantees that A_1 is false.[19]

---

[16] Vapor space is the space in a fuel tank occupied by fuel vapor and air rather than by the liquid fuel.

[17] Note that these further evidential questions are also evaluative questions concerning A_1: Neither of them presupposes that A_1 is true.

[18] As a reminder, a narrative answer is simply a (candidate) answer to a narrative question.

[19] This asymmetry does not hold in evaluative questions in general.

Not all evaluative questions about the narrative answer A_1 are binary questions like Q_1 and Q_2. For instance, questions Q_4 and Q_5 are also evaluative questions about A_1: Neither of these questions presupposes A_1, and the answers to these questions can provide evidence to evaluate A_1. For Question 4 ("What was the ratio of fuel molecules versus air molecules in the vapor space of the center wing tank at the time of the accident?"), we know that some fuel/air ratios support ignition, whereas others do not. Therefore, some candidate answers to Question 4 would provide evidence for A_1, whereas other candidate answers to Question 4 would provide evidence against A_1. Note also that the strength of evidential support varies from one candidate answer to another, because some fuel/air ratios may make ignition more likely than others. A similar analysis also applies to Question 5.

Third, conditional questions about the narrative answer A_1 provide a different kind of tests from evaluative questions about A_1. For an evaluative question about A_1, we obtain evidence for or against A_1 by determining *which* of the candidate answers to the evaluative question is correct. In contrast, for a conditional question about A_1, we obtain evidence for or against A_1 by determining *whether* the conditional question has a correct answer at all.

Consider Question 3 ("What was the ignition source inside the aircraft that ignited the flammable fuel/air mixture in the center wing tank and caused an explosion?"). This is a conditional question about A_1 because it presupposes A_1 (that some ignition source internal to the aircraft had ignited a flammable fuel/air mixture in the center wing tank and caused an explosion). If A_1 is true, then Question 3 should have *some* correct answer. If A_1 is false, however, Question 3 is not meaningfully posed and should not have a correct answer.

Consequently, we can obtain evidence for or against A_1 by determining whether Question 3 appears to have a correct answer. On the one hand, if we can obtain sufficiently strong evidence in support of one of the candidate answers to Question 3 and rule out the alternatives, that would count as evidence for A_1 no matter what the favored candidate answer turns out to be. On the other hand, if every candidate answer to Question 3 faces strong objections or if we cannot come up with any plausible answer to Question 3, that would count as evidence against A_1, because it would give us some reason to believe that A_1—a presupposition of Question 3—might not be true.

In the TWA 800 investigation, the investigators examined numerous possible ignition sources. Most of the candidate ignition sources could be ruled out easily. For instance, some possible ignition sources could not generate enough ignition energy to ignite a flammable fuel/air mixture; some possible ignition sources would leave behind characteristic traces that could not be found in the wreckage; there were no realistic mechanisms to transfer the ignition energy from some of the possible ignition sources to the center wing tank. If it had been the case that *every* candidate ignition source faced strong objections like these, the investigators would have needed to step back and reconsider whether A_1 was true at all.

Fortunately for the investigators, they were able to identify at least one plausible candidate ignition source, a short circuit event that transferred excess voltage from some high voltage wire to the fuel gauge wires. For this possible ignition source, the investigators were able to establish that (1) it could generate enough ignition energy to ignite a flammable fuel/air mixture in the center wing tank, (2) there were realistic transfer mechanisms for bringing the ignition energy into the center wing tank and releasing it there, and (3) there were electrical anomalies right before the accident that suggested that such a short circuit occurred. For these reasons, the investigators deemed the short circuit involving fuel gauge wires to be the most likely ignition source. They fell short of concluding that it was definitely the ignition source, because they could not rule out some other possible ignition sources with certainty. ([13], p.306-308) Nevertheless, such an inconclusive epistemic situation still counts as evidence in favor of A_1. This is because it suggests that there likely was an ignition source, even if we cannot say for sure what it was.

Fourth, despite the differences between evaluative questions and conditional questions about A_1, the two types of evidential questions do share a common characteristic: In the process of answering these evidential questions, if we find evidence in favor of A_1, we simultaneously discover additional details that enrich A_1 as a causal hypothesis. These additional details can then be incorporated into the narrative of the past event.

By itself, A_1 is not very detailed. It simply states that some ignition source internal to the aircraft had ignited a flammable fuel/air mixture inside the center wing tank and caused an explosion. It does not tell us what the ignition source was, and how the ignition and the explosion exactly happened.

Now, suppose the investigators evaluated A_1 by investigating Q_1 ("was there an inflammable fuel/air mixture?"), and suppose that they found evidence suggesting that the answer to Q_1 was positive. In the process of doing so, they should have identified a few additional details: What the temperature within the center wing fuel tank was, what the fuel/air molecule ration was, etc. If there were a fuel/air explosion, these additional details would be relevant to the narrative of such an explosion, because they would be informative about *how* such an explosion occurred. Similarly, if the investigators found evidence suggested that Q_3 ("what was the ignition source") has a plausible answer, they would also have identified additional details (what the ignition source was, how much ignition energy it could generate, how the energy was transferred to the center wing tank, etc) that were informative about how the explosion occurred.

In sum, both evaluative questions and conditional questions play a dual evidential role. For a promising candidate narrative answer, the evidential questions play the role of *narrative expansion*. That is, answering the evidential questions helps to expand the promising narrative answer into a fuller narrative. For the other less promising alternative answers, the evidential questions play the role of *cross checks*. That is, answering the evidential questions helps the investigators ensure that they have not overlooked anything and that the alternatives can indeed be ruled out. Such a dual evidential role contributes to the resolution of narrative questions.

To conclude, here is a compact summary of my account of evidential questions and how they contribute to the resolution of narrative questions:

- Let Q^N be a narrative question, and let $\mathcal{A}^{\mathcal{N}}_Q = \{A^N_1, \ldots, A^N_k\}$ be a set of candidate answers associated with Q^N at a time t during the investigation. For each candidate narrative answer A^N_i, the investigators can raise evidential questions *about it* at time t. Let Q^E be an evidential question about candidate answer A^N_i, and let $\mathcal{A}^{\mathcal{E}}_Q = \{A^E_1, \ldots, A^E_j\}$ be a set of candidate answers to the evidential question Q^E.
- If the evidential question Q^E is an *evaluative* question about A^N_i, then its set of candidate answers $\mathcal{A}^{\mathcal{E}}_Q = \{A^E_1, \ldots, A^E_j\}$ can be partitioned into two subsets $\mathcal{A}_1$ and $\mathcal{A}_2$. Every member of $\mathcal{A}_1$ provides some evidence for A^N_i, and every member of $\mathcal{A}_2$ provides some evidence against A^N_i.
- If the evidential question Q^E is a *conditional* question about A^N_i, then any plausible candidate answer in $\mathcal{A}^{\mathcal{E}}_Q$ provides at least some evidence for A^N_i. If there is a *unique* plausible candidate answer in $\mathcal{A}^{\mathcal{E}}_Q$, that provides even stronger evidence for A^N_i. In contrast, if every candidate answer in $\mathcal{A}^{\mathcal{E}}_Q$ faces strong objections, that constitutes evidence against A^N_i.
- A candidate answer A^N_i to the narrative question Q^N is *ruled out* at time t, if the total evidence against A^N_i at time t is sufficiently strong. To resolve a narrative question Q^N at time t—at least provisionally—is to possess a total body of evidence that supports one of the candidate answers to Q^N and rules out all the other candidate answers to Q^N at time t.

This completes my discussions of the resolution of questions in the context of event reconstruction research. In the next chapter, I discuss another important dimension of questions, namely the *significance* of questions.

Chapter 15

Question Significance

15.1 Bromberger On the Values of Questions

The questions that arise within event reconstruction research have different degrees of significance. If a relatively insignificant question was not resolved during the investigation, its lack of resolution tends to make little difference to event reconstruction research's overall success. In contrast, if a highly significant question was not resolved during the investigation, its lack of resolution may suggest that the investigation failed to accomplish what it was supposed to accomplish or that the investigation's main findings are suspect. In short, the significance of questions is a dimension of the investigative process that is relevant to the overall success of the event reconstruction research and the evidential quality of its results.

But what does it mean for a question to be more or less significant? What types of questions tend to be more significant in event reconstruction research? In this chapter, I propose an account of the significance of questions in event reconstruction research. The basic philosophical framework of my account is derived from Sylvain Bromberger's discussions of the values of questions. Even though Bromberger addressed a different philosophical question from mine, the concepts he introduced are helpful and applicable to questions that arise in event reconstruction research.

In his book *On What We Know We Don't Know*, Bromberger examines how an epistemic agent who wants to reduce her ignorance should make rational decisions about what questions to ask. More precisely, his philosophical question is this: If there is a set of questions that a rational epistemic agent P does not know the correct answers to at a given time t, and the current goal of this rational agent is to select the next question to be eliminated from P's ignorance, How can the agent rationally select the question? ([17], p.146)

Bromberger argues that for the agent to rationally investigate a question Q_1 in her pool of questions at a given time, that question Q_1 must satisfy four conditions. First, the agent must have the conceptual apparatus to understand and formulate Q_1. Second, the agent must know that Q_1 is sound; that is, it has a correct answer. For the agent to know this, she must know that the presuppositions of the questions are satisfied. Third, the agent must know that she does not yet have sufficient reason to accept any particular answer to Q_1 as being correct. Fourth, her choice of selecting Q_1 (or her policy of selection behind the choice) optimizes the balance of maximizing the values and minimizing the costs[1] of the selected question. ([17], p.148-151)

The main part of Bromberger's discussions that is of interest to me is his account of the *values* of a question. Bromberger distinguishes between four different types of values that a question can have:

1. The *Peircean value* of a question depends on the expected intellectual value of its answer, which is measured in terms of the "pleasure, relief or cognitive stability expected from its answer". ([17], p.152)

2. The *Jamesian value* of a question depends on the material benefits expected from its answer, for instance, the answer may enable us to construct something, to repair something, to find something, etc. ([17], p.152)

---

[1] Bromberger says very little about the costs of questions. Here is his one-sentence summary of the "many kinds of costs": "cost in time required for retrieval of information, for computation, for mastering value adders; financial costs of instrumentation, assistance; and emotional costs of boredom, anxiety or frustration." ("Rational Ignorance", in *On What Know We Don't Know*, p.142.)

3. The *Collingwood value* of a question depends on “the prospect that its answer will raise new questions with expected values of their own”. ([17], p.152)
4. The *Machian value* of a question depends on the “The prospect that its answer will yield answers to other questions with values of their own”. ([17], p.152)

Based on the wording of Bromberger’s definitions, it seems that all four types of values are *expected* values of questions at a given time. This suggests the following interpretation of Bromberger’s definitions: First, the primary bearers of values are candidate answers rather than questions, and the value of a question depends on which candidate answer is correct. Second, to estimate the expected value of a question at a given time, the rational agent needs to assume (1) a provisional set of candidate answers to that question, (2) a value for each candidate answer at that time, and (3) how likely each candidate answer is expected to turn out to be the correct answer at that time.

If the above interpretation correctly reflects Bromberger’s view, then Bromberger’s notion of the *value* of a question is not the same as my notion of the *significance* of a question. In my view, the significance of a question is not equivalent to the expected value of its answer. Moreover, Bromberger’s account of the values of questions is also too abstract to be sufficiently informative about questions that arise in event reconstruction research.

Nevertheless, I find Bromberger’s distinction between the four kinds of values of questions to be instructive, because it helps me to make a similar distinction about the significance of questions. In the rest of this chapter, I distinguish between four kinds of question significance, parallel with Bromberger’s distinction: Peircean significance, Jamesian significance, Machian significance, and Collingwood significance. I examine each type of question significance in turn, illustrating it with examples drawn from case studies of NTSB investigations.

15.2 Peircean Significance

Bromberger defines the Peircean value of a question as the expected intellectual value of its answer. Correspondingly, I define the Peircean significance of a question as the intellectual significance of that question.

What *is* the intellectual significance of a question? It depends on the inquiry in which the question arises. Every inquiry has certain *goals* that need to be accomplished, and some of these goals are *intellectual* goals. A question that arises in an inquiry is intellectually significant just in case answering it helps to fulfill some intellectual goals of the inquiry. The *degree* of the intellectual significance of a question depends on the extent to which answering the question helps fulfill some intellectual goal of the inquiry, and on the intellectual significance of the goal in that inquiry.[2]

In short, the Peircean or intellectual significance of a question is derived from the question's relationship to the intellectual goals in an inquiry. What endows Peircean significance on a question is the *fact* that answering the question fulfills certain intellectual goals that are important in the inquiry. A question can have Peircean significance at a given time in an inquiry, even if the investigators have not yet come up with any candidate answer to that question at the time. Moreover, the Peircean significance of the question does not depend on which candidate answer to the question turns out to be correct. Neither does not require a probability distribution on the candidate answers or a value assignment to these answers. Therefore, the Peircean significance of a question is not equivalent to the expected intellectual value of its answer.

To understand the Peircean significance of questions that arise in *event reconstruction research*, we need to understand the intellectual goals of event reconstruction research. An event reconstruction research is, by definition, an investigation that reconstructs a past event. The intellectual goals of an event reconstruction research can be divided into two

---

[2] I take the (intellectual) significance of an intellectual goal in an inquiry to be a primitive notion.

main types: The *idiographic* intellectual goals, and the *nomothetic* intellectual goals.

The terms "idiographic" and "nomothetic" were introduced by the Kantian philosopher Wilhelm Windelband. Windelband distinguishes between what he calls "idiographic" knowledge and "nomothetic" knowledge. In his view, idiographic knowledge concerns "what once was"; it is time and culture-bound and is essentially historical. In contrast, nomothetic knowledge concerns "what always is", namely immutable, ahistorical, general laws.[3]

The idiographic versus nomothetic distinction has been generalized by researchers to apply to two types of research methodologies in history, political science, and social sciences. The idiographic method is characterized by scenario-driven reasoning: it focuses on an in-depth understanding of individual cases and their unique histories, and it often resists generalizing over cases. In contrast, the nomothetic method is characterized by generalization-driven reasoning: It focuses on understanding individual cases as members of populations or classes, and it often relies on theoretical generalizations and laws (sometimes quantitative theories and laws) for explaining individual cases.

I believe that the idiographic versus nomothetic distinction can also be applied to characterize the intellectual *goals* of event reconstruction research. For instance, one of the main intellectual goals of an event reconstruction research is to be able to construct a *narrative* of the past event in question. This intellectual goal is idiographic, because it focuses on understanding what happened during an *individual* event, why it happened and how it happened the way it did. The narrative of the past event typically includes case-specific and unique elements, and the narrative as a whole is typically not generalizable to other events. For instance, a narrative of the French Revolution would almost certainly include elements unique to the French Revolution and not generalizable to other revolutions.

In contrast, another main intellectual goal of event reconstruction research is to understand the type-level causal systems that are at work in the event in question. This

---

[3] See James Lamiell, "'Nomothetic' and 'idiographic': Contrasting Windelband's Understanding with Contemporary Usage", [26].

intellectual goal is a nomothetic goal, because it focuses on understanding the individual event in question as a member of an event type and the type-level causal regularities that can help explain the individual event. For instance, a nomothetic intellectual goal of a study of the French Revolution could be to identify type-level causal structures shared by other revolutions such as the Russian Revolution and the Chinese Revolution.

A question that arises in an event reconstruction research has Peircean significance, if answering the questions helps fulfill either an idiographic intellectual goal or a nomothetic intellectual goal. For instance, why-questions, how-questions, and wh-questions about anomalies (including abnormal traces and abnormal subevents) in a complex event typically have high degrees of Peircean significance, because answering these questions helps identify what is distinctive or unique about the event in question, and the answers to these questions are incorporated into the narrative of the complex event.

Of course, not all the questions about anomalies in a complex event have an equal amount of Peircean significance. Not all the anomalies are equally abnormal, and not all the anomalies play an equally important role in the complex event. In the American Airlines 191 investigation, for instance, the separation of the No. 1 (left) engine was arguably the most important anomaly in the accident. This is partly because it was an extremely unusual event, and partly because it was causally connected with numerous other anomalies that contributed to the crash. Therefore, the question "why did the No.1 engine fall off?" arguably has the highest degree on Peircean significance in this investigation.

Next, questions about type-level causal systems that could be at work in a complex event tend to have a high degree of Peircean significance, partly because answering these questions helps satisfy the nomothetic intellectual goals of the investigation. For instance, during the TWA 800 investigation, the investigators established that an overpressure event had occurred in the center wing tank, which suggested that a fuel/air explosion in the center wing tank might have occurred. They soon realized, however, that much of the basic information about the flammability of Jet A fuel vapor and its combustion behavior was not available. Most of the existing flammability research focused on single-component fuels. In contrast, Jet A fuel had a complex chemical composition, and the center wing tank also had a complex geometry. ([14])

As a result, the NTSB investigators carried out a series of flight tests and laboratory experiments to determine relevant causal parameters such as the flashpoint of Jet A fuel and the peak pressure generated by the ignition of the Jet A fuel inside the center wing tank. The questions about the Jet A fuel ignition mechanism have Peircean significance in this investigation, for two reasons: First, the answers to these questions help determine whether a fuel/air explosion in the center wing tank had occurred in the accident, which in turn shapes the narrative of the accident. Second, answering the questions helps the investigators understand how the Jet A fuel could ignite and generate explosive pressure in the center wing tanks of the entire fleet of Boeing 747-100 aircraft, of which TWA flight 800 was a member.

15.3 Jamesian Significance

Bromberger defines the Jamesian *value* of a question as "the value that a question gets from the material benefits expected from its answer". ([17], p.152) I define the corresponding notion of the Jamesian *significance* of a question somewhat differently, as the practical significance of that question.[4] The term "practical" contrasts with the term " intellectual, "intellectual", and I use these two terms more or less according to their common-sense meaning.

Just as the Peircean significance of a question depends on the *intellectual* goals of the inquiry, the Jamesian significance of a question depends on the *practical* goals of the inquiry. Moreover, the distinction between the idiographic and the nomothetic also applies to the practical goals of an inquiry. The idiographic goals of the inquiry are those directed towards

---

[4] Practical concerns can go beyond material needs and benefits. For instance, the attribution of moral responsibility is a practical concern.

a particular case, event, or individual, whereas the nomothetic goals of the inquiry are those directed towards a type of cases, events or individuals.

Different inquiries can have very different practical goals; this is true even among different types of event reconstruction research. In this section, I will focus on the practical goals of *engineering failure investigations.* First, a fundamental practical goal of engineering failure investigations that is *idiographic* is to attribute responsibility and assign blame or praise to various parties implicated in the failure. Sometimes, an engineering failure investigation also provides an analysis of the failure that could be used in litigation.

Second, a fundamental practical goal of engineering failure investigations that is *nomothetic* is to improve the safety of the relevant engineering system and learn lessons that can help prevent similar failures from happening again in the future. In practice, this involves using an investigation as an opportunity to discover potential weakness and vulnerabilities in the design, manufacture, maintenance, and operation of an engineering system; in the selection, training, and support of the personnel interacting with the system; and in the organizational culture in which the engineering system is situated. Often, the investigation can discover vulnerabilities that may or may not be related to the accident in question, long before the investigations determined why and how the accident happened; these vulnerabilities also have to be fixed. The NTSB, for instance, often started issuing safety recommendations before it completed an investigation.

In engineering failure investigations, questions with high Jamesian significance are those most critical for fulfilling the above fundamental idiographic and nomothetic goals. In particular, counterfactual questions concerning how the engineering failure could have turned out differently have high Jamesian values, because the correct answers to these questions can be useful for either the idiographic or the nomothetic goals of the investigation.

Since the motivations behind the counterfactual questions can be either idiographic or nomothetic, we can distinguish between two types of counterfactual questions in engineering failure investigations. The first type of counterfactual question is idiographic: They focus on what could or could not have happened *in this particular case.* For instance, suppose a given outcome happened. Could it have turned out differently, if so, how? Questions concerning the necessity or contingency of an outcome are often idiographic. For instance, at what point in the sequence of events could the outcome still have been changed, and at what point did it become inevitable? These counterfactuals are crucial for fulfilling the idiographic practical goals of the investigation, such as attributing responsibility, blame, and praise.

The second type of counterfactual question is nomothetic: They focus on what could or could not have happened in other cases similar to this case. For instance, given that this event happened to this aircraft, could it have happened to other aircraft of the same type? Given that this pilot made such an error in this situation, would other pilots have made the same error in similar situations? These counterfactual questions are crucial for fulfilling the nomothetic practical goals of the investigation, such as identifying and eliminating safety hazards and vulnerabilities in a type of engineering systems.

The following hypothetical example provides a clear illustration of the distinction between idiographic and nomothetic counterfactual questions.[5] A flight crew preparing for takeoff failed to set the flaps to the takeoff position. Flaps are control surfaces on the trailing edge of wings used to provide extra lift for takeoff. As a result of the pilots' failure to set the flaps, the airplane could not generate enough lift while taking off, and it stalled and crashed shortly after. Setting the flaps is a highly practiced procedure step that the crew should have performed thousands of times before, which gives rise to the following questions: "Why did the crew fail to set the flaps? Although the crew did not set the flaps, could they have done any differently in this case?"

Both the why question and the counterfactual question above have an idiographic focus. The point of asking these questions is to understand what had happened, and what could

---

[5] The example is taken from *The Limits of Expertise: Rethinking Pilot Error and the Causes of Airline Accidents*, by Dismukes, Berman, and Loukopoulos. [20], p.3.

have happened, in *this very* case. The counterfactual question, in particular, is needed for assessing responsibility and blame.

Unfortunately, the idiographic counterfactual question can be difficult to answer, because it may not be possible to determine why the crew overlooked setting the flaps. Even if the crew members survived the crash and are interviewed, they may not be able to explain the oversight. The NTSB may be able to determine various factors such as fatigue, distraction that *might* have contributed to the oversight. Still, it can be difficult to determine to what extent these factors contributed to the error in this particular case.

Instead of asking the idiographic counterfactual question, the NTSB instead asks a related but different, nomothetic counterfactual question:

> If a population of pilots with experience and skills comparable to those of the accident crew faced a situation similar to the situation faced by the accident crew, would this population of pilots be vulnerable to making the same kinds of errors made by the accident crew? If so, why? ([20], p.3)

I will make three comments about this nomothetic counterfactual question. First, the focus of this question is *not* what actually happened to the specific crew in a specific accident and what could have happened differently in that very situation. Rather, the focus is on whether a *type* of vulnerability—the vulnerability of all similarly skilled pilots to the same kind of error in similar situations—exists. The purpose of asking this question is to come up with insights that can help prevent similar accidents from happening again.

Second, another way to see the nomothetic nature of the above question is to notice that it is a probability question (with a frequentist interpretation). In asking "would a population of pilots similar to the accident crew be vulnerable to making an error similar to this particular error?", the investigators are looking for the percentage of a population of pilots making that error. Occasional errors made by pilots and other skilled experts occur in a somewhat random fashion, which is why the nomothetic counterfactual question should be interpreted probabilistically: At a population level, it is more appropriate to speak of factors influencing the probability of error rather than deterministically causing the error.

Finally, in answering this nomothetic counterfactual question, we imagine a counterfactual scenario that only preserves some of the accident's characteristics. Which common characteristics of the accident are held constant in the counterfactual scenario depends on the practical interests of the investigators. In human error investigations, recent approaches tend to emphasize the role of systematic factors that enable human errors to occur and to produce catastrophic consequences; of particular interest are organizational processes that can create hazardous local working conditions and defeat safety redundancies.[6] Given these specific interests, the relevant organizational structures and their deficiencies will be held constant in the nomothetic counterfactual scenario, even though the particular pilots and some details of the flights may be changed.

In sum, questions motivated by the fundamental practical goals of an inquiry have higher Jamesian significance. In the context of engineering failure investigations, the questions with the highest Jamesian significance tend to be idiographic counterfactual questions that serve the goal of responsibility attribution, or nomothetic counterfactual questions that serve the goal of safety improvement.

15.4 Machian Significance

Bromberger defines the Machian value of a question as "the value that a question gets from the prospect that its answer will yield answers to other questions with values of their own"([17], p.152).[7] According to Bromberger, Machian values depend on devices—e.g., laws and theories—that turn answers to some questions into answers to other questions. ([17], p.152-153) For instance, the equation $T \approx 2\pi\sqrt{L/g}$ transforms the answers to the

---

[6] For instance, certain airlines may be negligent in providing detailed information on takeoff from contaminated runways; Operational control that is supposed to provide crucial support to flight crews may be poorly trained and ill-qualified; etc.

[7] Bromberger says that he names this type of value "Machian" because Mach valued laws and theories, and Bromberger's account of Machian values relies on such laws and theories (which he called "value adders").

question "what is the length of the pendulum?" into answers to the question "what is the period of the pendulum?", assuming that the amplitude is small.

My notion of the Machian *significance* of a question shares a similar idea: A question Q has Machian significance in an inquiry, just in case answering Q helps answer other questions in that inquiry. I call those questions that have Machian significance **contributory questions**, because they contribute to answering other questions. However, there are many possible ways in which answering one question helps answering another question. The type of scenario that Bromberger describes, namely that each candidate answer to one question is transformed into a candidate answer to another question via an equation, law, or theory, is just *one* of the possible ways in which a question can contribute to another. In this section, I describe a few other types of contributory questions in event reconstruction research, i.e., a few other ways in which questions in this type of inquiry possess Machian significance.

First, we have already seen a type of contributory questions in the previous chapter, namely *evidential questions* about candidate answers to another question. More precisely, let Q^N be a narrative question that arises in an event reconstruction research, and let $\mathcal{A}_Q^N = \{A_1^N, \ldots, A_k^N\}$ be a set of candidate answers associated with Q^N at time t during the investigation. For each candidate narrative answer $A_i^N \in \mathcal{A}_Q^N$, we can raise evidential questions that serve as tests of this candidate answer. Some evidential questions are evaluative questions; others are conditional questions. Either way, how these evidential questions are answered provides evidence either for or against candidate answer A_i^N. I claim that these evidential questions are also contributory questions for Q^N, because answering each evidential question helps to evaluate a particular candidate answer to Q^N, which in turn contributes to the resolution of Q^N.

Second, a question P can be a contributory question for another question Q, in the sense that (1) the answer to P implies a partial answer to Q, and (2) the partial answer, when combined with additional information, can be converted into a full answer to Q. For instance, in the TWA 800 investigation, one of the key questions in the early phases of the investigation was: "What was the inflight breakup sequence?" This question (call it Q) asks for a step-by-step account of how the aircraft broke up in the air, which involves arranging all the fractures and separation of aircraft components in a temporal sequence. It is a substantial narrative question whose answer adds a lot of details to the narrative of the accident. Fully answering it was a challenging task that required the collaboration of multiple NTSB subgroups.

One of the NTSB subgroups approached question Q by answering an easier question P: "What was the spatial distribution of the wreckage of TWA flight 800?" Answering question P contributes to answering question Q, because the spatial distribution of the wreckage reflects, approximately and at a very coarse-grained level, which *portions* of the aircraft separated earlier, and which *portions* separated later. For instance, by identifying the distribution of wreckage underwater[8], the investigators identified three main areas of wreckage, which they called the red, yellow, and green zones, respectively. The red zone is closest to the JFK airport (the departing airport) along the airplane's flight path, while the green zone is located furthest from JFK along the flight path. Hence, the investigators inferred that the red zone components were the first to separate, followed by the yellow zone components, and finally the by green zone components. ([13], p.69-73)

In short, question P is a contributory question for question Q in this example, in the sense that the answer to P implies a partial or coarse-grained answer A_p to Q. The partial answer A_p, however, is not detailed enough to count as a breakup sequence. For instance, the red zone pieces include the forward portion of the wing center section and a ring of fuselage directly in front. Knowing that these were the earliest pieces to separate from the airplane still does not tell us which of these pieces broke up first. To convert the partial answer A_p into a full answer to question Q, more information is needed. Another NTSB subgroup searched for the additional information by doing a three-dimensional reconstruction of the

---

[8] TWA 800 departed from the JFK airport and broke up in the air over the Atlantic Ocean near East Moriches, New York.

recovered wreckage, examining each of the three portions of the structure in great detail, and developing localized sequences of failures based on observed features of fractures in each portion. ([13], p.103) When the investigators combined the coarse-grained sequence with the localized sequences, they arrived at a full answer to the question Q: “What was the breakup sequence?”

Third, a question Q_1 can be a *potentially* contributory to another question Q_2, if the following two conditions hold: (1) there exist some candidate answers to Q_1 that imply (or support) some candidate answers to Q_2, (2) there exist some candidate answers to Q_1 that are irrelevant to Q_2. In other words, depending on what the correct answer to Q_1 turns out to be, answering Q_1 may or may not help answer Q_2, and the only way to know for sure is to answer Q_1 first and find out whether its answer is helpful. In this case, we may say that the question Q_1 has *potential* Machian significance. Such a question Q_1 can still be worth trying out, especially if the question Q_2 is difficult to answer by other means.

One type of question in event reconstruction research that can have potential Machian significance are questions about abnormal subevents or abnormal traces. I argued earlier that questions of this type typically have high Peircean significance since answering these questions helps fulfill the intellectual goals of the investigations by contributing to the narrative of the complex event. Here, I further argue that at least some questions of this type have potential Machian significance, because answering a question about one anomaly in a complex event can potentially contribute to answering another question about a different anomaly in the same complex event.

Let E be a complex event under the investigation, N_1 and N_2 be two anomalies (abnormal subevents or traces) that occurred or existed within the event E. Further, let Q_1 be the question “Why did N_1 occur/exist?” and Q_2 be the question “Why did N_2 occur/exist?” Under *some* appropriate conditions, we may have good (but still defeasible) reasons to believe that the answers to Q_1 and Q_2 are related. For instance, we may have some reasons to believe that N_1 and N_2 had a common cause. Assuming that Q_2 is more difficult to answer than Q_1, we may consider treating Q_1 as a potentially contributory question for Q_2. However, it may still turn out that the answers to Q_1 and Q_2 are not related. To know whether the answer to Q_1 is relevant to Q_2 at all, we need to answer Q_1 first.

Consider an example from the USAir 427 investigation. Previously, I mentioned that one of the anomalies in this accident was that the aircraft began to yaw and roll to the left at a high rate starting from 19:02:58, and sustained the high rate of yaw and roll until the aircraft stalled. Call this anomaly N_2, and let Q_2 be the question "Why did the yaw and roll moment *initiate*?"

Moreover, another anomaly in the accident was that about the time when the yaw and roll initiated (about 19:02:57), the CVR recorded a sequence of three mysterious thumping sounds in 1 second. In the first few days after the crash, members of the CVR team listened to these sounds hundreds of times but could not recognize them. Call this anomaly N_1, and let Q_1 be the question "What was the source of the thumping sounds?"

The investigators thought it was important to answer the question Q_1 "What was the source of the thumps?", because the thumps occurred at a crucial time, right before the airplane started to yaw and roll left. ([1], p.136) It *could* be the case that the thumps and the initiation of the yaw and roll movement had a common cause, in which case the answer to Q_1 implies an answer to Q_2. However, it *could* also be the case that the thumps and the initiation of the yaw and roll movement were unrelated despite their temporal coincidence. In short, Q_1 was a potentially contributory question for Q_2. To determine whether answering Q_1 helps answer Q_2, the investigators had to answer Q_1 first and see if its answer supports any candidate answer of Q_2.

Using a combination of spectral analysis, acoustic fingerprinting, and comparisons with sounds recorded in flight tests, the investigators eventually arrived at an answer to Q_1, namely that the thumps were caused by USAir 427's encounter with wake vortices generated from a preceding aircraft ([10], p.131). Moreover, one of the candidate answers to Q_2 *also* happened to be USAir 427's encounter with the wake vortices from Delta flight

1083, and investigators already had some preliminary evidence for this candidate answer to Q_2: For instance, flight simulations showed that the wake vortices produced by airplanes comparable to Delta flight 1083 *could* have initiated the drastic yawing and rolling movement experienced by USAir flight 427.[9] ([18], p.130) The answer to Q_1 establishes that USAir 427 *did* encounter the wake vortices from Delta 1083; which, together with other available evidence, helps the investigators establish the correct answer to Q_2, namely that the encounter with wake vortices initiated the yaw and roll of USAir 427.

In sum, there are many different types of contributory questions (of which we have only surveyed a few). So there are many possible ways for a question to possess Machian significance.

15.5 Collingwood Significance

Finally, Bromberger defines the Collingwood value of a question as "the value that a question gets from the prospect that its answer will raise new questions with expected values of their own" ([17], p.152). Out of the four types of values of a question that Bromberger distinguishes, the notion of Collingwood value is arguably the most difficult to understand. For instance, what type of questions have the property that their answers tend to raise other valuable questions? Bromberger says little about the notion of Collingwood value and does not illustrate it with examples, and it is not obvious what intuitive idea this notion is supposed to capture.

Before formally introducing my notion of Collingwood *significance* of a question, I would like to motivate it with a more intuitive idea. Let Q be a question that arises in an event reconstruction research, and the research aims to construct a narrative N of a complex event E. The basic intuition behind my idea is the following: Q has a high degree of Collingwood significance, just in case different candidate answers to Q make a *substantial*

---

[9] The flight simulation also showed, however, that wake turbulence by itself was insufficient to *sustain* the yaw and roll for long, thus some other causes were needed to sustain the yaw and roll movement.

difference to the overall content of the narrative N. To put it differently, a question Q with a high degree of Collingwood significance is a question that is important to resolve correctly, for the following reason: If the question Q is resolved incorrectly, then, by the end of the event reconstruction research, a substantial part of the narrative N produced by the research would be incorrect as well.

What types of questions in event reconstruction research tend to have higher Collingwood significance in this intuitive sense? An inaccurate but suggestive proposal is that questions that arise early in the investigation tend to have higher Collingwood significance. The proposal is inaccurate because not all questions that arise early in the investigation make a substantial difference to the narrative's content. However, some early questions make a substantial difference, because when these questions were resolved, their answers both contributed to the narrative and became presuppositions of subsequent questions. Later, these subsequent questions will be resolved, and their answers also both contributed to the narrative and became presuppositions of further questions. If the early questions had been resolved incorrectly, then many of the subsequent questions would be based on false presuppositions, which implies that a substantial part of the narrative would be suspect.

So far, I have introduced an intuitive notion of Collingwood significance. To make the intuitive notion more precise, I introduce the following terminologies and definitions:

- Let P and Q be questions that have already been raised by time t in an investigation.
- Definition 1: P **directly depends on** Q at time t, just in case (1) Q is resolved at time t and has an answer A, and (2) A is a presupposition of P; i.e., if A were false, P would not be a meaningful question. Alternatively (to use graph terminologies), we call P a **child** of Q, and Q a **parent** of P.
- Definition 2: P **depends** on Q at time t, just in case there exists a sequence of questions $Q_1, \ldots, Q_n$ that have been raised by time t in the investigation, such that

(1) $P = Q_n$, (2) $Q = Q_1$, (3) Q_{i+1} directly depends on Q_i for each $1 \leq i \leq n-1$. To use graph terminologies, we call P a **descendant** of Q, and Q an **ancestor** of P.

- Definition 3 (First definition of Collingwood significance): The **Collingwood significance** of a question Q is the total number of questions that depend on Q at the end of the investigation.
- Definition 4 (Second definition of Collingwood significance): The **Collingwood significance** of a question Q is the combined significance[10] of all the questions that depend on Q at the end of the investigation.

I would like to make a few comments about the above definitions. First, Definition 3 and Definition 4 are two *nonequivalent* definitions of Collingwood significance, and I include both of them here because I believe that both of them capture useful ideas.

Definition 3 is a formalization that *approximates* the intuitive notion of Collingwood significance introduced earlier. It defines the Collingwood significance of a question Q as the total *number* of descendant questions[11] of Q at the end of the investigation, which is a proxy of the impact of question Q on the narrative obtained at the end of the investigation. After all, the more descendant questions Q has at the end of the investigation, the more consequential it is to resolve Q correctly; if Q had been resolved incorrectly, then its descendant questions would have been based on false presuppositions, and the answers to these descendant questions that have been incorporated into the narrative would all be suspect.

In contrast, Definition 4 takes into consideration the differences in the significance of descendant questions. It defines the Collingwood significance of a question Q as the combined significance of all the descendant questions of Q at the end of the investigation. Unlike Definition 3, Definition 4 prioritizes descendant questions that are more significant, in the

---

[10] Here the notion of significance encompasses all the four dimensions of significance. However, what matters the most is probably Peircean significance and Jamesian significance.

[11] To reiterate, saying that a question P is a descendant of a question Q means the same thing as saying that P depends on Q, as defined on the previous page.

sense that more significant descendant questions of Q contribute more to the Collingwood significance of Q. It is an alternative approximation of the intuitive notion of Collingwood significance introduced earlier, emphasizing the fact that some descendant questions—e.g., those with high Peircean and Jamesian significance—are more central to the narrative of the complex event than other questions.

Next, even though the two definitions of Collingwood significance are not equivalent, I believe that they are fairly well correlated. A question that has high Collingwood significance according to one definition tends to have high Collingwood significance according to the other definition. Consequently, I will not make a judgment about which of these two definitions is “correct”. For the rest of this dissertation, I will use the term “Collingwood significance” with Definition 4 in mind.

Finally, note that the Collingwood significance of a question is only defined at the end of the investigation. Intuitively, this is because there is no way to determine the full impact of a resolved question on the entire narrative of the complex event until the end of the investigation. Of course, we can still *estimate* the Collingwood significance of a question before the investigation ends, using heuristics based on what type of question it is and how early it arises in the investigation.

To illustrate the terminologies and definitions I have introduced so far, consider three questions from the TWA 800 investigation. First, let Q_1 be the question: “What initiated the inflight breakup sequence?” Answering Q_1 was one of the major tasks in the early phases of the TWA 800 investigation. Based on the wreckage distribution, the investigators determined that the wing center section was the first portion of the aircraft that separated. By examining a 3-dimensional reconstruction of the wreckage of wing center section, and tracing the directions of crack growth on the wreckage, the investigators established that the first component of the aircraft that failed was the center wing fuel tank inside the wing center section. Finally, the witness marks and deformations found in the center wing fuel tank indicated that the fuel tank broke apart because of an overpressure event inside it. At this point, the investigators resolved question Q_1. They concluded that its correct answer was A_1: "The inflight breakup of the airplane was initiated by an overpressure event in the center wing tank."

Second, based on A_1, the investigators raised a further question Q_2: "Why did such an overpressure event (that initiated the inflight breakup) occur in the center wing tank?"[12] They considered two main candidate answers to Q_2: Either a high energy detonation device such as a missile or a bomb detonated near the center wing tank, or some internal aircraft component provided ignition energy that ignited the fuel/air vapor inside the tank. After recovering over 95 percent of the fuselage and not finding any characteristic signs of high energy explosion, the investigators ruled out the possibility of high energy detonation, resolved question Q_2 and concluded that its correct answer was A_2: "Some ignition source internal to the aircraft had ignited a flammable fuel/air mixture inside the center wing tank and caused an explosion."

Third, based on A_2, the investigators raised yet another question Q_3: "What was the ignition source that ignited the flammable fuel/air mixture and caused an explosion?"[13] After considering numerous candidate answers to Q_3 and ruling out most of them, the investigators eventually concluded that the most likely ignition source was a short circuit involving the fuel gauge wires. However, they cautioned that they were not certain about this conclusion because they could not rule out a few alternative ignition sources definitively. In other words, the investigators did not *quite* resolve Q_3 in the end, and they were only able to conclude that its most likely answer was A_3: "The ignition source was a short circuit that transferred excess energy from some high voltage wire to the fuel gauge wiring."

Consider the relationship between Q_1, Q_2 and Q_3 at the end of the TWA 800 investigation. According to my definitions:

---

[12] To clarify, the investigators did not have to wait until Q_1 was resolved before raising Q_2. As long as A_1 was a candidate answer to Q_1 and was not ruled out, the investigators could raise Q_2 based on A_1.

[13] In the previous chapter, I argued that Q_3 is an *evidential* question about A_2, because the existence of plausible answers to Q_3 provides additional evidence for the candidate answer A_2. Again, Q_3 could be raised as long as A_2 remained a candidate answer to Q_2. Hence Q_3 could be raised even before Q_2 was resolved.

- At the end of the investigation, Q_2 directly depends on Q_1. This is because (1) Q_1 is resolved at the end of the investigation and has answer A_1, and (2) A_1 is a presupposition of Q_2.

- At the end of the investigation, Q_3 directly depends on Q_2. This is because (1) Q_2 is resolved at the end of the investigation and has answer A_2, and (2) A_2 is a presupposition of Q_3.

- At the end of the investigation, both Q_2 and Q_3 depend on Q_1.

- By Definition 3 or Definition 4, Q_1 has the highest Collingwood significance, Q_2 the second-highest, the Q_3 the lowest Collingwood significance. This follows from the fact that Q_2 is a descendant question of Q_1 and Q_3 is a descendant question of Q_2.

Moreover, we can see intuitively *why* Q_1 has a high Collingwood significance, by engaging in the following thought experiment:

At the end of the TWA 800 investigation, since both Q_1 and Q_2 were resolved, their answers A_1 and A_2 were incorporated into the narrative of the accident; even the (most likely) answer A_3 to question Q_3 could be incorporated into the narrative, as long as a "most likely" qualifier was attached to it.

Let us imagine, however, that the investigators later realized that they had resolved Q_1 incorrectly. For instance, perhaps newly emerged evidence showed definitively that the initial failure of the aircraft was not the overpressure of the center wing fuel tank, but rather the rapid separation of a cargo door. Moreover, what the investigators thought were traces of an overpressure event in the center wing tank were actually just fractures and deformations caused by separations of other parts of the aircraft. In short, it was not the case that an overpressure event in the center wing tank initiated the breakup sequence, and the candidate answer A_1 to question Q_1 turned out to be false.

But if A_1 was false, then it would no longer be meaningful to pose question Q_2. If no overpressure event in the center wing tank initiated the rest of the breakup sequence, it would no longer make sense to ask what caused such an overpressure event. Moreover, the meaningfulness of question Q_3, which presupposes an answer to Q_2, would become questionable (at the very least). Consequently, the investigators would have to remove all three answers A_1, A_2, and A_3 from the narrative of the accident. They could no longer state any of the following: A short circuit involving fuel gauge wires likely ignited the fuel/air vapor in the center wing tank, which in turn caused an overpressure event in the center wing tank, which in turn initiated the rest of the breakup sequence.

In sum, if Q_1 turned out to have been resolved incorrectly, then a substantial portion of the narrative of the accident would have to be revised. In contrast, incorrect resolution of Q_2 or Q_3 would not require as substantial a revision of the narrative of the accident. Intuitively, this is why Q_1 has the highest Collingwood significance among the three questions.

This completes my discussions of the four types of the significance of a question in event reconstruction research. My main ideas can be summarized as follows: First, the Peircean significance of a question consists of the fact that answering it helps fulfill certain intellectual goals of the investigation. Second, the Jamesian significance of a question consists of the fact that answering it helps fulfill certain practical goals of the investigation. Third, the Machian significance of a question consists of the fact (or at least the potential) that answering it helps to answer other significant questions in the investigation. Finally, the Collingwood significance of question consists of the fact that many other significant questions depend on it at the end of the investigation.

In the next chapter, I analyze the last major component of question dynamics—how questions arise, and propose an account of the coherence of narratives based on the full structure of question dynamics.

Chapter 16

Question Dynamics and Coherence

16.1 The Arising of Questions

In the previous chapters, I have used expressions such as “question Q arises” and “X gives rise to question Q” liberally. These expressions capture the idea that an inquiry is a dynamic process where new questions continue to emerge, and answers to our questions open up further questions. However, what do these expressions *mean*? In this chapter, I begin by proposing an account of what it means for a question to arise, and what it means for something to give rise to a question. Such an account of the arising[1] of questions, together with my account of the resolution and the significance of questions in the previous chapters, allow me to describe the full question-and-answer process of an idealized inquiry, which I call *question dynamics*.[2] Finally, based on the question dynamics of event reconstruction research, I propose my account of the *coherence* of narratives of past events.

Let me begin with two preliminary clarifications about the way I use the phrase “question Q arises”. First, to say that a question arises in an investigation is *not* to say that it was actually *raised* by investigators in that investigation. Which questions are actually

---

[1] The verb “arise” is a technical term on my account, and I use “arising” as the noun version of “arise”.

[2] I borrowed the term “question dynamics” from Nicholas Rescher. See [30], chapter 4.

asked and puzzled over in an investigation can depend on many contingent factors that are not relevant for my purpose. Instead, what I need is a notion of the arising of questions that is useful for a rational reconstruction of the question-and-answer structure of event reconstruction research.

Second, to say that a question arises is to describe a *state*, not an *event.* A question can be in one of the two states at any time t during an investigation: Either it arises at time t, or it does not arise at time t. A question may first arise at time t_1, and then continue to arise after t_1.[3] Moreover, I use the phrase "question Q *emerges*" to denote a transition from the state in which Q does not arise to the state in which Q arises, and I use the phrase "question Q *dissolves*" to denote a transition from the state in which Q arises to the state in which Q does not.

What does the statement "a question Q arises" mean, then? In the existing literature on questions, the statement "a question Q arises" is typically interpreted as "Q can be meaningfully posed". For instance, according to Nicholas Rescher, "a question *arises* at time t if it then can meaningfully be posed because all of its presuppositions are then taken to be true." ([30], p.46)

Similarly, the existing literature typically interprets "a declarative statement X gives rise to a question Q" as expressing a semantic relation, which is usually characterized as "X implies that the question Q has a correct answer" or "X implies all the presuppositions of Q", with some variations in details. For instance, in a discussion of Erotetic Logic, Sylvain Bromberger writes: "Roughly put, a proposition *gives rise* to a question if it entails that the question has a correct answer. A proposition is a *presupposition* of a question if its falsehood entails that the question has no correct answer... Notice that if a proposition gives rise to a question, then it entails the presupposition of that question. The converse does not hold." ([17], p.120)

Similarly, Andrzej Wisniewski defines the *evocation* of a question by a set of declarative

---

[3] Arguably, this is not how the phrase "a question arises" is used in ordinary speech.

formulas as follows: ([39], p.12)

> A question Q is evoked by a set of declarative formulas X just in case all the following conditions hold:
>
> (a) No direct answer to Q belongs to X.
>
> (b) No direct answer to Q is entailed by X.
>
> (c) If all the formulas in X are true, the question Q must have a true direct answer.
>
> (d) Each presupposition of Q is entailed by X.

To paraphrase Wisniewski's account: a set of propositions gives rise to a question just in case it implies all the presuppositions of the question, guarantees that the question has a correct answer, but does not entail any particular answer to the question.

In short, authors such as Rescher, Bromberger, and Wisniewski subscribe to a semantic account of the arising of questions. On such an account, a question Q arises just in case it is meaningfully posable, and a declarative statement S (or a set of declarative statements) gives rise to a question Q just in case S implies that Q is meaningfully posable. In turn, the meaningful posability of a question can be characterized as "the question is guaranteed to have an answer" or "all the presuppositions of the questions are true", etc.

However, I argue that the semantic account of the arising of questions is inadequate for my purpose. To be clear, the notion of the meaningful posability of a question is not problematic in its own right. In fact, it is a useful notion that I have already relied on twice in the previous chapters: For instance, both the notion of a conditional question and the notion of direct dependence involves one question Q_2 presupposing an answer A_1 to another question Q_1, which implies that Q_2 is not meaningfully posable if A_1 is ruled out.[4]

---

[4] The notion of conditional question and the notion of direct dependence are essentially the same notion used in different contexts. The notion of conditional question is primarily used in a context when we are evaluating candidate answers to an open question Q_1. In this context, a conditional question is a question Q_2 that presupposes one of the candidate answers to the open question Q_1, and how Q_2 is answered provides evidence for or against Q_1. In contrast, the notion of direct dependence is primarily used in a context when we are determining the Collingwood significance of a resolved question. In this context, a question Q_2 directly depends on a question Q_1, just in case Q_1 is resolved and Q_2 presupposes the answer to Q_1.

However, the semantic account of the arising of questions does not adequately capture the notion that I am interested in. My goal is to understand the question-and-answer structure of an (idealized) investigation. I need an account of the arising of questions that explains why some questions but not others *should* belong to the investigative agenda at a given point during the investigation. The semantic account is far too permissive for this purpose: At any time in an investigation, there are innumerably many questions that can be meaningfully posed at that time, the vast majority of which are completely irrelevant to the investigation and do not belong to the investigative agenda. Consequently, I need an alternative account of the arising of questions.

The semantic account does provide a necessary condition for (my notion of) the arising of questions: If a question Q is not meaningfully posable at time t—for instance, if some of its presuppositions are ruled out at time t—then Q definitively should not belong to the investigative agenda at time t. What other necessary conditions are needed for the arise of questions? Intuitively, a question arises at time t in an investigation, only if it is *worth answering* for the investigators at time t. To put it another way, another necessary condition is that the question has at least some degree of *significance* for the investigation at time t.

I believe that the two necessary conditions—meaningful posability and a degree of significance for the investigation—are jointly sufficient for a question to belong to the investigative agenda. In other words, I propose the following account of the arising of questions:

> A question Q arises at time t in an investigation, if and only if:
>
> - Q can be meaningfully posed at time t. That is, no presupposition of Q is ruled out in the investigation at time t.
> - Q possesses at least some degree of significance for the investigation at time t. Here "significance" can be either Peircean significance, or Jamesian significance, or Machian significance.

I have two comments about this account of the arising of questions. First, I define "Q can be meaningfully posed at time t" in a specific way, namely that the presuppositions of Q are not ruled out at time t. My point is that to meaningfully pose a question Q at time t during an investigation, the investigators do not need to believe that the presuppositions of Q are true. Rather, they only need to believe that the presuppositions of Q *could be true* given what they already know, i.e., the presuppositions of Q have not yet been ruled out by what they know at time t.

Consider a familiar example from the TWA 800 investigation. Let Q_1 be the question "Why did an overpressure event occur in the center wing fuel tank?". This question Q_1 was open at some point during the investigation (call it time t) with two main candidate answers: A_1 is the candidate answer "a high energy detonation device detonated near the center wing tank and caused an explosion", whereas A_2 is another candidate answer "an ignition source internal to the airplane ignited the fuel/air vapor inside the tank and caused an explosion". At time t, the investigators could (and did) raise evidential questions about each of these candidate answers. Moreover, one type of evidential questions is what I called the *conditional* questions, which presuppose a given candidate answer.

For instance, concerning candidate answer A_1, the investigators could raise the following conditional questions: "What was the high energy detonation device that detonated near the center wing tank? If the high energy detonation device was a missile, where was the missile launched from? If the high energy detonation device was a bomb, where was the bomb located?" All of these questions presuppose the candidate answer A_1, and all of them were meaningful to pose at time t. However, the investigators did not regard candidate answer A_1 as true at time t. Instead, they regarded A_1 as being compatible with what they already knew at time t, and the point of asking the conditional questions was to evaluate A_1: If there do not exist plausible candidate answers to the conditional questions, that counts as evidence against A_1; conversely, if there exist plausible candidate answers to the conditional questions, that counts as evidence for A_1.

Similarly, concerning candidate answer A_2, the investigators could raise conditional questions such as: "What was the ignition source that ignited the inflammable fuel/air vapor in the center wing tank? How much ignition energy did the ignition source supply, and how was the ignition energy transferred from the ignition source to the center wing tank?" Again, all of these questions presuppose the candidate answer A_2; all of them were meaningful to pose at time t, and yet the investigators did not regard candidate answer A_2 as being true at time t either. Instead, they only regarded A_2 as not yet ruled out based on existing knowledge, and used the conditional questions to evaluate A_2.

My second comment is that, according to my account of the arising of questions, the statement "every question that arises within an investigation has at least some degree of significance for the investigation" is a tautology, because "having at least some degree of significance for an investigation" is part of what "arises in an investigation" means. Moreover, there are three main ways in which a question can be significant for an investigation: (1) answering it could fulfill some intellectual goals of the investigation, (2) answering it could fulfill some practical goals of the investigation, and (3) answering it could help to answer other questions that are significant for the investigation. In short, Peircean, Jamesian, and Machian significance are all relevant to the arise of questions. The outlier is Collingwood significance, which is only defined at the end of the investigation and does not apply to the arising of questions in general.[5]

Next, I propose an account of the meaning of the expression "X gives rise to a question Q" that goes beyond the semantic account. On my account, X could be many different things in the context of event reconstruction research:

- X could be declarative statements; for instance, declarative statements that describe abnormal subevents or traces give rise to why and how questions, and declarative statements that describe candidate answers to an open question give rise to evidential

---

[5] Collingwood significance is essentially a notion of the significance of *correct resolution* of questions.

questions that evaluate the candidate answers.

- X could also be questions; for instance, a question P can give rise to a contributory question Q, and answering Q contributes to answering P.
- X could even be a narrative or partial narrative of the complex event; for instance, a narrative of the event can give rise to counterfactual questions about whether the event could have happened differently; and if so, how.

I will say more about what I mean by a narrative later in the chapter; for now, think of a narrative as the final product of an event reconstruction research, and a partial narrative as a partially constructed narrative during the investigation. Narratives and partial narratives are complex cognitive products that can give rise to a far greater range of questions than individual declarative statements or questions can. Counterfactual questions are just one type of questions that narratives can give rise to.

Despite the variety of things that could give rise to questions, my account of the meaning of the expression "X gives rise to a question Q" remains the same no matter what X is. Here is my account:

> X gives rise to a question Q at time t in an investigation, if and only if:
>
> - Q can be meaningfully posed at time t. That is, no presupposition of Q is ruled out in the investigation at time t.
> - There exists some facts F about X at time t, such that F endow Q at least some degree of significance for the investigation at time t.[6] Here "significance" can be either Peircean significance, or Jamesian significance, or Machian significance.

Let me use three examples to illustrate this account. First, let X be a declarative statement that describes an abnormal subevent E. Moreover, X gives rise to a why-question Q "Why

---

[6] To put it differently: There exists some facts F about X at time t, such that F make Q at least somewhat significant for the investigation at time t.

did E occur?" at time t. On my account, to say that X gives rise to Q at time t is to say the following: (1) the presupposition of Q—the occurrence of the subevent E—is not ruled out at time t; (2) there exists some facts F about X that endow Q some degree of significance for the investigation at time t. Here, the fact F is that the subevent described by X is abnormal, and F endows Peircean significance on the question Q.

To make the example more concrete, let E be the simultaneous failure of all three hydraulic systems in the United 232 accident. In this case, condition (1) says that the presupposition of the why-question Q—the simultaneous failure of all the three hydraulic systems—is a possibility that the investigators could not rule out and had to seriously entertain. This is definitely the case in the United 232 example, where the investigators were nearly certain that such a triple hydraulic failure had occurred when they raised the why-question Q.

Moreover, condition (2) says that the fact that the subevent described by the declarative statement X was abnormal endows the question Q at least some degree of Peircean significance. This condition also holds in the United 232 example, since a triple hydraulic failure in an aircraft was so unusual that the investigators needed to understand why and how such a failure occurred; no narrative of the accident could be complete without making sense of this failure.

Second, let X be a question that arises in the investigation, and X gives rise to a contributory question Q. On my account, to say that X gives rise to Q at time t is to say the following: (1) the presupposition of Q is not ruled out at time t, and (2) there exists some facts F about X that endow Q some degree of significance for the investigation at time t. Here, the facts F include the fact that X is a question that has some degree of significance for the investigation, and the fact that answering Q helps (or at least potentially helps) with answering X. These facts endow Machian significance on the question Q.

Finally, let X be a fully or partially constructed narrative of an engineering accident, and X gives rise to a counterfactual question Q. Again, On my account, to say that X gives rise to Q at time t is to say the following: (1) the presupposition of Q is not ruled out at time t, and (2) there exists some facts F about X that endow Q some degree of significance for the investigation at time t. Here, the facts F include the fact that part of the purpose of having a narrative of the accident is to fulfill the practical goals of the investigation, which include idiographic goals such as responsibility attribution, and nomothetic goals such as preventing similar future accidents. Moreover, F also include the fact that responsibility attribution and future accident prevention require answering appropriate counterfactual questions. These facts endow Jamesian significance on the question Q.

To illustrate this last example, let X be an entire narrative of the history of flight of United 232, including a detailed account of how the pilots crash-landed the aircraft at Sioux City airport. This narrative gave rise to two counterfactual questions: First, could the pilots have safely landed the aircraft on the runway without any hydraulic control? Second, if a population of pilots was specifically trained for landing without hydraulic flight controls, how likely would it be for them to land the airplane successfully in a similar situation? The narrative gave rise to the first counterfactual question, because part of the purpose of this narrative is to fulfill the idiographic goal of determining whether the actions of the pilots are blameworthy/praiseworthy, which requires answering the first counterfactual question. The narrative gave rise to the second counterfactual question, because part of the purpose of this narrative is to fulfill the nomothetic goal of determining whether training for triple hydraulic failure would help to improve the success of landing in the future, which requires answering the second counterfactual question.

So far, I have given an account of what it means for questions to arise, and an account of what it means for something to give rise to questions. To reiterate, the arising of a question is a state that can continue after it first obtains: a question can arise at a given time and continue to arise afterward, unless it ceases to arise at a later point. Similarly, something can give rise to a new question at a given time, and continue to give rise to that question afterward. If we want to specifically refer to the first moment when a question arises, we can use expressions such as: “a question Q first arises at time t”, or “a question Q emerges at time t”, or“X gives rise to a new question at time t”. If we want to refer to the first moment when a question goes from the state of “arising” to the state of “not-arising”, we can use expressions such as “the question Q ceases to arise at time t”, or “the question Q dissolves at time t”.

This completes my account of the arising of questions. In the next section, I will use all the conceptual resources that I have introduced so far—the arising, the resolution, and the significance of questions—to describe the full structure of question dynamics in event reconstruction research.

16.2 The Structure of Question Dynamics

A question dynamics is the process of question-and-answer in an idealized inquiry. It is a dynamic process because new questions continue to arise even as old questions are resolved. It is a rational reconstruction because it is based on an idealized inquiry; for instance, it is primarily concerned with what questions *should* belong to the investigative agenda at a particular point in the investigation, rather than with what questions were actually investigated. In the previous chapters and the previous section of this chapter, I have examined the main components of questions dynamics in event reconstruction research by studying a few actual NTSB investigations, which I assume to approximate idealized investigations because of their successes. This section aims to put everything together and make explicit the full structure of question dynamics in event reconstruction research.

We can conceptualize the question dynamics of an event reconstruction research as being divisible into a series of discrete time steps. It has a beginning and an end. The following parameters characterize each time step of the question dynamics:

- A set $\mathcal{O}$ of questions that arise at time t and are still open at time t.

- A set $\mathcal{R}$ of questions that arise at time t and have been (provisionally) resolved at time t.
- Every question Q in the set $\mathcal{O}$ or in the set $\mathcal{R}$ is associated with a degree of significance at time t.
- Every question Q in the set $\mathcal{O}$ or in the set $\mathcal{R}$ is associated with a set $\mathcal{A}$ of candidate answers to Q at time t.
- Every candidate answer A to a question Q in the set $\mathcal{O}$ or in the set $\mathcal{R}$ is associated with a collection of available evidence for or against A at time t.
- Every question Q in the set $\mathcal{R}$ is associated with a degree of the provisionality of resolution at time t.
- A set $\mathcal{S}$ of all the answers to the (at least provisionally resolved) questions in $\mathcal{R}$ at time t.

Here is a list of definitions of (and comments on) the key terms used to describe the above parameters, which also shows how different parameters are related to each other:

- A question Q is *open* at time t, just in case the corresponding set $\mathcal{A}$ of candidate answers is not a singleton at time t. That is, $\mathcal{A}$ is either an empty set or a multi-set.
- A question Q is *resolved* at time t, just in case the corresponding set $\mathcal{A}$ of candidate answers is a singleton at time t.
- The degree of significance associated with each question that arises at time t is a measure that combines all four dimensions of significance—Peircean, Jamesian, Machian, and Collingwood significance—of the question at time t.
- A candidate answer A to a question Q at time t is a viable full answer to Q at time t. That is, it is an answer that, for all that the investigators know at time t, could turn out to be the correct answer to Q.

- The collection of available evidence for or against a candidate answer A at time t includes how the evidential questions about A are answered at time t.

- For a resolved question Q, its degree of the provisionality of resolution is a measure of how provisional its resolution is at time t. It depends on the strength of evidence for the one remaining candidate answer to question Q, and the extent to which the evidence can rule out possible alternative answers to Q.

Moreover, every single parameter of question dynamics can change over time. This includes the set $\mathcal{A}$ of candidate answers *and* the degree of significance associated with each question Q that still arises (open or resolved). The set of candidate answers can change over time for two reasons: First, certain candidate answers that used to be in the set $\mathcal{A}$ can be ruled out later. This happens when there is sufficient evidence against these answers, such that they are no longer considered viable and are thus excluded from the set $\mathcal{A}$. Second, certain new candidate answers that have not been considered before may be added to the set $\mathcal{A}$ at time t.

The degree of significance associated with a question can also change over time. For example, suppose a candidate answer A to a question P gives rise to a question Q, perhaps because A describes an anomaly, which endows Peircean significance on Q. If the evidence for the candidate answer A decreases over time, however, the significance of Q will also decrease over time. If A is ruled out eventually, then Q may cease to arise altogether. For another example, suppose at one point Q has Machian significance because it is deemed necessary for answering another question P. If the investigators later find another way of answering P without having to address Q, that reduces the Machian significance of Q as well.

The question dynamics of an idealized inquiry over time is a recursive process. First, the **base case** of the recursion concerns how new questions arise at the beginning of the investigation. In the context of event reconstruction research:

- Routine questions arise automatically at the beginning of the investigation, and they go into set $\mathcal{O}$.
- Any immediate observations of anomalies (abnormal traces or subevents) give rise to questions, which go into set $\mathcal{O}$.

For instance, in aviation failure investigations, there is a collection of routinely asked questions at the beginning of nearly every plane crash. These questions are generic and presuppose very little about the facts of a particular case, and therefore apply to a wide range of cases of the same type. For instance, if a crash is involved, the investigators would ask: "How is the wreckage distributed?" If an in-flight breakup is involved, the investigators would ask: "Which pieces of the aircraft separated first, and which pieces separated later? What was the temporal sequence of the separations?" If a component fracture is involved and the fracture surfaces are available, the investigators would ask: "What type of fracture is this?"

Moreover, a routine part of an aviation failure investigation is to conduct a *comprehensive search for anomalies* at the beginning of the investigation, and the questions that arise in this search are also routine. For instance, after a plane crash, some investigators are tasked with examining the accident aircraft's engines and engine accessories. The routine questions that they would ask include: "Were the engines functioning normally before the crash? Alternatively, were there any signs of pre-impact engine malfunction that could have contributed to the crash?"

Similarly, other investigators would be tasked with examining other aspects of the accident, including the airframe structure, the electrical and hydraulic systems, the air traffic control communications, the personal and medical background of the pilots, the maintenance records of the accident aircraft, and the weather conditions at the time of the accident, etc. The routine questions for all these investigators share the same form: "Were there any anomalies in aspect X of the accident that could have played a role?" Most of these routine anomaly-searching questions would not lead to any useful clues. However, some of them would, and the anomalies that they help identify would, in turn, give rise to further questions.

Next, the **recursion step** of the question dynamics applies to every subsequent time step of the investigation. It is a process in which new questions arise while (some of the) old questions are resolved at the same time:

- Questions in set $\mathcal{O}$ may give rise to new questions at time t, which will go into $\mathcal{O}$.
- For each open question, each of its candidate answers may give rise to new questions at time t, which will go into $\mathcal{O}$.
- Some previously open questions in $\mathcal{O}$ may be resolved at time t, so they will be removed from $\mathcal{O}$ and added to $\mathcal{R}$, and the answers to these questions will be added to $\mathcal{S}$.
- Some previously resolved questions in $\mathcal{R}$ may get reopened at time t, so they will be removed from $\mathcal{R}$ and added back to $\mathcal{O}$, and the answers to these questions will be removed from $\mathcal{S}$.
- If a candidate answer A to a question in either $\mathcal{O}$ or $\mathcal{R}$ is ruled out at time t, then all the questions that presuppose A cease to arise at t. Moreover, all the questions that A gave rise to ceases to arise at t.
- Some questions in $\mathcal{O}$ or $\mathcal{R}$ may cease to arise at time t. If a question Q ceases to arise at time t, then it is removed from either $\mathcal{O}$ or $\mathcal{R}$. If question Q was in $\mathcal{R}$, its answer is removed from the set $\mathcal{S}$. Moreover, any other question that presupposes any candidate answer to Q also ceases to arise at time t.
- The set $\mathcal{S}$ of answers to (provisionally) resolved questions at time t can give rise to new questions, and there are two possibilities. If $\mathcal{S}$ already contains enough information

to resolve these questions, then they go to $\mathcal{R}$. Otherwise, they go to $\mathcal{O}$.

Note that ruling out a candidate answer to a question can have a rippling effect on the question dynamics, because it nullifies all the questions that are "downstream" of that candidate answer.

Finally, the investigation will **terminate** at the final time point. Sometimes, the investigation terminates for epistemic reasons: For instance, perhaps all the available information has been used, and no new question arises. Other times, the investigation terminates for pragmatic reasons: Perhaps it has to end due to lack of funding, an approaching deadline, etc. When the investigation terminates, the set $\mathcal{O}$ of open questions may be nonempty, and the output of the investigation is the final set $\mathcal{S}$ of answers to resolved questions in $\mathcal{R}$. As we shall see, A comparison between the two sets $\mathcal{O}$ and $\mathcal{R}$ at the end—how many questions are still in each set, what are the degrees of significance associated with the questions in each set, etc.—can be indicative of the evidential status of the final output $\mathcal{S}$ of the investigation.

Depending on the nature of the inquiry, $\mathcal{S}$ may be converted into different cognitive products. For instance, in the context of a theoretical inquiry, $\mathcal{S}$ is converted into a theory. In the context of event reconstruction, the questions that arise in the investigation are *narrative questions*; $\mathcal{S}$ is a set of *narrative answers*; and $\mathcal{S}$ is converted into a *narrative* of a complex past event.

16.3 Coherence of Narratives

In this section, I propose an account of the *coherence* of narratives of past events, based on the structure of question dynamics in event reconstruction research.

Before laying out my account of coherence, I will clarify what I mean by a narrative of a past event. A narrative of a past event is an account of the causal structure of that event. It details what happened in the event, how it happened, and why it happened the way it did. I mentioned earlier that the set $\mathcal{S}$ of answers to resolved questions at the end of an event reconstruction research could be *converted* into a narrative of a past event. Now, I want to emphasize that the narrative is not *equivalent* to the set $\mathcal{S}$ of answers to resolved questions; rather, the narrative is *based on* the set of answers.

My main argument for distinguishing between narratives and the collections of answers on which they are based is that narratives have aesthetic and rhetorical qualities that go beyond the collection of answers produced by the question dynamics. What makes a narrative *good* often goes beyond the *correctness* of the answers on which the narrative is based. Two writers who agree on the collection of answers produced by the investigation may nevertheless construct very different narratives.

For instance, one rhetorical device that narratives typically deploy is *focalization*, which refers to the point of view through which we see events in the narrative unfold. The story of a plane crash can be told from the perspective of the flying crew, that of the passengers, that of the investigative agencies, that of the aviation industry, that of an omniscient being, etc. A narrative may also shift from one perspective to another within the storytelling. Focalizing can contribute richly to how we think and feel as we read the narrative, but it is also a highly value-laden device that can color our perception of the event.

Perhaps because of the aesthetic and rhetorical qualities inherent in the narratives of past events, the NTSB—one of the most successful agencies for reconstructing failure events—does not construct narratives in their accident reports. If we go through a typical NTSB report, nowhere does it contain a full narrative of the failure event in question. Instead, the NTSB report provides detailed answers to questions that arise in the investigation and supports the answers with the evidence that the investigators have collected. In turn, the answers contained in the NTSB report put others (e.g., news reporters and popular science book writers) in a position to construct narratives of the failure events in question.

In short, the distinction between narratives of past events and the collection of answers on which the narratives are based reflects good methodology. Focusing on the collection of answers on which narratives of past events are based helps us disentangle the evidential considerations from the aesthetic considerations, making it easier to make progress on the former.

Finally, let $\mathcal{O}$, $\mathcal{R}$, and $\mathcal{S}$ be the set of open questions, the set of resolved questions, and the set of answers to the resolved questions respectively, all obtained at the end of event reconstruction. Let N be a narrative of a past event based on the answers in $\mathcal{S}$. The **coherence** of the narrative N is a measure of the overall evidential status of $\mathcal{S}$; that is, how well the answers in $\mathcal{S}$ are supported by the total evidence obtained in the investigation. Coherence comes in degrees, and the degree of coherence of the narrative N can be characterized by the following parameters:

- The size of the set $\mathcal{O}$ of open questions at the end of the investigation.
- The size of the set $\mathcal{R}$ of resolved questions at the end of the investigation.
- The degree of significance of each question in $\mathcal{O}$.
- The degree of significance of each question in $\mathcal{R}$.
- The degree of the provisionality of resolution for each question in $\mathcal{R}$.

Moreover, the degree of coherence of the narrative N depends on the values of the parameters in the following way:

- The greater the size of $\mathcal{R}$ is, the higher the degree of coherence is. That is, resolving more questions that arise in the investigation increases the coherence of the narrative.
- The greater the size of $\mathcal{O}$ is, the lower the degree of coherence is. That is, failures to resolve questions that arise in the investigation decrease the coherence of the narrative.
- The higher the degree of significance of any question in $\mathcal{R}$ is, the higher the degree of coherence is. That is, resolving questions of greater significance that arise in the investigation increases the coherence of the narrative.

- The higher the degree of significance of any question in $\mathcal{O}$ is, the lower the degree of coherence is. That is, failures to resolve questions of greater significance that arise in the investigation decrease the coherence of the narrative.
- The less provisional the resolution of each question in $\mathcal{R}$ is, the higher the degree of coherence is. That is, more convincingly resolving questions that arise in the investigation increases the coherence of the narrative.
- The greater the *negative correlation* between the degree of significance of each question in $\mathcal{R}$ (on the one hand) and its degree of the provisionality of resolution (on the other hand), the higher the degree of coherence is. That is, more convincingly resolving questions of greater significance increases the coherence of the narrative.

This completes my account of the coherence of narratives of past events. To gain more intuition about this account, consider some corollaries that follow from it:

- Corollary 1: A narrative of a past event is more coherent, if why-questions, how-questions, and wh-questions that arise from anomalies (abnormal traces or subevents) are satisfactorily resolved (i.e., their resolutions are less provisional).

Corollary 1 follows from the fact that questions that arise from anomalies tend to have high degrees of Peircean significance. This corollary captures the intuition that a (more) coherent narrative of a past event should account for as many of the abnormal traces as possible, and it should incorporate as many known abnormal subevents as possible. To use a slogan, the ideal of a coherent narrative is “no anomaly left out”. Of course, this slogan only applies to anomalies that (the investigators believe) are related to the event being reconstructed. Moreover, the ideal typically is not realized in practice. Nevertheless, to produce a minimally coherent narrative, investigators should *at least* strive to satisfactorily resolve questions that arise from *highly* abnormal subevents or traces, such as the engine separation in the AA 191 accident and the triple hydraulic failure in the United 232 accident.

Some may object to Corollary 1 with the following argument: It is not clear that satisfactorily resolving a question that arises from an anomaly increases the coherence of the narrative. For instance, suppose we just resolved a why-question Q_1 that arises from an anomaly A_1. But the answer to Q_1 presumably describes another anomaly A_2, which gives rise to another why-question Q_2. In turn, resolving Q_2 presumably gives us an answer that describes another anomaly A_3, which gives rise to yet another why-question Q_3. Every time we resolve a why-question that arises from an anomaly, we simultaneously create at least one new open why-question that arises from another anomaly. We can keep doing this by tracing the causal chain backward. Therefore, trying to explain as many anomalies as possible in a narrative is futile, because it only leads to our discovering more and more anomalies that are unexplained.

In response, I argue that the objection makes the following assumption: That an abnormal subevent E_1 tends to be caused by a comparably abnormal subevent E_2, which in turn tends to be caused by a comparably abnormal subevent E_3, and so forth. I reject this assumption. In fact, a highly abnormal subevent E, such as an engine separation, a fuel tank explosion, or a rudder reversal, tends to be caused by the co-occurrence of numerous subevents and conditions, virtually all of which are either perfectly normal or only slightly abnormal. Part of what explains the abnormality of E is simply the fact that the co-occurrence of all of its causes is incredibly unlikely. Yet the co-occurrence happened, by chance, which does not need to be explained. In short, I believe that the process of explaining abnormal subevents in a complex past event will reach a point where there is no reason to trace the causal chain backward any further.

- Corollary 2: A narrative of a past event is more coherent, if questions with high Collingwood significance are satisfactorily resolved (i.e., their resolutions are less provisional).

Recall that the Collingwood significance of a question is the combined significance of all the questions that depend on it at the end of the investigation. Intuitively, questions with high Collingwood significance are very important to resolve correctly; because if they were resolved incorrectly, a substantial part of the narrative would be incorrect and would need to be revised. For instance, questions about the breakup sequence of an inflight breakup have high Collingwood significance, because if the investigators got the breakup sequence wrong, they would likely get everything about the causes of the inflight breakup wrong.

Consequently, Corollary 2 captures the intuition that narratives with high degrees of coherence are less vulnerable to large scale revisions in the future. Moreover, the protection against large scale future revisions is not magical, it is derived from facts about the structure of the question dynamics that help create the narrative. To use a metaphor, we can think of a more coherent narrative as corresponding to a question dynamics whose "key joints" are reinforced, such that the overall structure of the question dynamics is less vulnerable to destruction by the pressure of new evidence that emerges in the future.

- Corollary 3: A narrative of a past event is less coherent, if there exist open questions that are p-predicaments at the end of the investigation.

The term "p-predicament" is due to Sylvain Bromberger, who defines it as follows: "Someone is in a p-predicament with regard to some question Q, if and only if on that person's views, the question Q admits of a right answer, can generate from his mental repertoire no answer to which given that persons' views, there are no decisive objections." ([17], p.81) Here, I re-appropriate the term "p-predicament" and define it in a different way. According to my definition, a question Q is a p-predicament at time t, if and only if (1) it presupposes the answer A to a resolved question P, (2) the set of candidate answers $\mathcal{A}$ associated with Q is an empty set at time t, i.e., all the candidate answers to question Q have been ruled out at time t.

P-predicaments tend to be *very* bad for the coherence of a narrative, for the following reasons: First, a p-predicament is an open question in itself, which decreases the coherence of the narrative. But (much) more importantly, a p-predicament significantly increases the likelihood that its presupposition—which is an answer to a resolved question—is actually false. For instance, if question Q presupposes an answer A to a resolved question P, then the fact that Q has no viable candidate answer at all significantly raises the probability that A is false. This will likely force the question P to reopen, which reduces coherence. Moreover, if A is sufficiently in doubt that it is ruled out, all the other downstream questions that presuppose A will be nullified as well. In short, P-predicaments have great potential to cause large scale revisions of a narrative.

Corollary 3, then, captures the intuition that more coherent narratives are less vulnerable to large scale future revisions for another reason—they typically do not have p-predicaments. To use a related metaphor, we can think of a more coherent narrative as corresponding to a question dynamics that does not have "weak joints". That is, the question dynamics lacks the kinds of questions that have the potential to bring down a large part of its structure when under the pressure of new evidence that emerges in the future.

Let me conclude this chapter by returning to the jigsaw puzzle metaphor that motivated my account of coherence in the introduction to Part III. There, I suggested that a coherent narrative is analogous to a completed jigsaw puzzle where "all the pieces fit together", and one way to think about what "all the pieces fit together" means is to think about how we solve jigsaw puzzles sequentially. To produce a jigsaw puzzle where "all pieces fit together", we just need to add one piece at a time, use the earlier pieces as constraints for later ones, and continually add pieces given the constraints of the previous pieces. Similarly, to create a coherent narrative, we just need to resolve one question at a time, use the answers to earlier questions as constraints for later questions, and continually resolve new questions given the constraints provided by the answers to the earlier questions.

My account of coherence in this chapter captures this intuitive conception of coherence. According to my account, the more questions that arise in the question dynamics are resolved, the more significant and satisfactorily resolved these questions are, the more coherent the narrative is. However, my account also goes beyond this basic intuition, by showing how the coherence of narratives depends crucially on the *structure* of the question dynamics that produces them. For instance, it shows how more coherent narratives are less vulnerable to radical future revisions, due to structural facts about their questions dynamics. Of course, the account of coherence I have given in this chapter is far from complete. But it does suggest that a question-oriented approach to the evidential status of narratives of past events is not just useful, but *essential*.

Chapter 17

Conclusion to Part III

17.1 Recap

Let us recap the major findings of the chapters in this part of the dissertation.

Chapter 14 ("Question Resolution") discusses the resolution of questions in the context of event reconstruction research. I began by introducing an abstract account of what it means to resolve a question in an investigation. On this account, at a given time t within an investigation, any question Q that has risen by time t is associated with a set $\mathcal{A}_Q$ of full candidate answers to it. The question Q is *open* at time t, just in case the set $\mathcal{A}_Q$ is either empty or has multiple elements at time t. The question Q is *resolved* at time t, just in case the corresponding set $\mathcal{A}_Q$ is a singleton at time t. To resolve a question Q is to possess a body of evidence that supports one of the candidate answers in $\mathcal{A}_Q$ and rules out all the alternative candidate answers, so that $\mathcal{A}_Q$ becomes a singleton. The resolution of a question is more or less provisional, depending on how strong the evidence in support of the one remaining answer is, and the extent to which the evidence is sufficient to rule out possible alternative answers.

To situate this abstract account of question resolution in the context of event reconstruction research, I further examined the types of questions that arise in event reconstruction research, what constitute candidate answers to these types of questions, and two forms of evidence for or against these candidate answers in Chapter 14.

First, since an event reconstruction research ultimately enables the investigators to construct a *narrative* of a complex past event, the answers to many (if not all) questions that arise in the event reconstruction research must have contributed to the narrative. A question Q that arises in an event reconstruction research is a *narrative question*, just in case if Q is resolved, its answer will be incorporated into the narrative of the event. Each candidate answer to a narrative question is a *narrative answer*; and narrative answers are typically statements about the causal structure of the event being reconstructed. Consequently, narrative questions are *queries* about the causal structure of the event, and they typically take the form of why-questions, how-questions, and wh-questions such as what, when and where-questions.

Second, I further examined what constitutes candidate answers to the main forms of narrative questions. Since what count as candidate answers to wh-questions are relatively straightforward—for instance, a candidate answer to a where-question takes the form of a location—I focused on the forms of answers to why-questions and how-questions. With regard to why-questions, I endorsed a version of van Fraassen's contrastive theory, and argued that a candidate answer to a why-question in event reconstruction research describes a purported *difference maker*—i.e., a causal factor that purportedly makes the difference between the topic of the why question and a specific set of contrasts. With regard to how-questions, I argued that a candidate answer to a how-question describes causal processes that purportedly produced an outcome. By "causal processes that produce X", I mean (1) a collection $\mathcal{C}$ of events and conditions that, taken together, are causally sufficient for X to occur; and (2) the sequences of intermediate events leading from the initial events and conditions in $\mathcal{C}$ to the outcome X.

Finally, I examined how investigators obtain evidence for or against candidate answers to a narrative question. I proposed that the investigators typically obtain evidence by raising and answering further questions—which I called "evidential questions"—for each candidate answer. The structures of the evidential questions are informative about the forms of evidence for or against a candidate answer. Moreover, I distinguished between two main types of evidential questions, the *evaluative* questions and the *conditional* questions, respectively.

If an evidential question Q about a candidate answer A is an evaluative question, then it does not presuppose that candidate answer A. Moreover, the set of candidate answers to Q can be partitioned into two subsets $\mathcal{A}_1$ and $\mathcal{A}_2$. Every member of $\mathcal{A}_1$ provides some evidence for A, and every member of $\mathcal{A}_2$ provides some evidence against A. In contrast, if an evidential question Q about a candidate answer A is a conditional question, then Q presupposes A. Moreover, the existence of at least one plausible candidate answer to Q provides some evidence for A, whereas the lack of any plausible candidate answer to Q provides some evidence against A.

Chapter 15 ("Question Significance") discusses the significance of questions in event reconstruction research. Inspired by Bromberger's distinction between four types of values of questions in an inquiry, I made a similar distinction between four types of the significance of questions in event reconstruction research: Peircean significance, Jamesian significance, Machian significance, and Collingwood significance.

First, a question that arises in an event reconstruction research has Peircean significance, just in case answering it helps fulfill some intellectual goals of the research. An event reconstruction research typically has two types of intellectual goals: *Idiographic* intellectual goals directed towards understanding the past event in question as a unique individual event, and *nomothetic* intellectual goals directed towards understanding the past event in question as a member of an event type. For instance, constructing a full narrative of the past event is typically an idiographic intellectual goal, whereas understanding type-level causal structures that are at work in the event is a nomothetic intellectual goal. Why-questions and how-questions that arise from anomalies occurred in the event tend to fulfill either idiographic or nomothetic intellectual goals. Hence they tend to have high Peircean significance in event reconstruction research.

Second, a question that arises in an event reconstruction research has Jamesian significance, just in case answering it helps fulfill some practical goals of the research. Since different types of event reconstruction research tend to have very different practical goals, I focus on a particular type of event reconstruction research, namely engineering failure investigations. The practical goals of engineering failure investigations can be further divided into idiographic practical goals and nomothetic practical goals. The idiographic practical goals of engineering failure investigations include assigning blame, praise, and responsibility to the parties involved. In contrast, the nomothetic practical goals include safety improvement and prevention of similar failures in the future. To fulfill either the idiographic or the nomothetic practical goals, engineering failure investigators often need to answer *counterfactual* questions. Hence these questions tend to have high Jamesian significance in engineering failure investigations.

Third, a question that arises in an event reconstruction research has Machian significance, just in case answering it contributes to (or at least has the potential of contributing to) answering other questions with significance of their own; hence I called questions that have Machian significance *contributory questions.* There are many different types of contributory questions, and I considered a few: For instance, evidential questions are contributory questions, because they provide evidence for or against a particular candidate to a narrative question, which helps with resolving that narrative question. Moreover, a question can be contributory to another question, if the answer to the first is a partial answer or coarse-grained answer to the second. Finally, a question can be *potentially* contributory to another question, if some candidate answers to the former are relevant to resolving the latter, while other candidate answers to the former are irrelevant to the latter. A why-question about an anomaly can potentially contribute to another why-question about another anomaly, if there is some reason to believe that the two anomalies had a common cause.

Finally, I motivated the concept of Collingwood significance with an intuitive idea: Intuitively, a question that arises in an event reconstruction research has Collingwood significance, just in case if the question had been resolved incorrectly, then a significant portion of the narrative of the past event would be incorrect as well. To make the concept of Collingwood significance more precise, I introduced the following definitions:

- Let P and Q be questions that have already been raised by time t in an investigation.
- Definition 1: P **directly depends on** Q at time t, just in case (1) Q is resolved at time t and has an answer A, and (2) A is a presupposition of P; i.e., if A were false, P would not be a meaningful question.
- Definition 2: P **depends** on Q at time t, just in case there exists a sequence of questions $Q_1, \ldots, Q_n$ that have been raised by time t in the investigation, such that (1) $P = Q_n$, (2) $Q = Q_1$, (3) Q_{i+1} directly depends on Q_i for each $1 \leq i \leq n-1$.
- Definition 3 (First definition of Collingwood significance): The **Collingwood significance** of a question Q is the total number of questions that depend on Q at the end of the investigation.
- Definition 4 (Second definition of Collingwood significance): The **Collingwood significance** of a question Q is the combined significance of all the questions that depend on Q at the end of the investigation.

I argued that both Definition 3 and Definition 4 approximate the intuitive idea of Collingwood significance, even though they are not equivalent. I assume Definition 4 as the default definition of Collingwood significance in the rest of the dissertation.

Chapter 16 ("Question Dynamics and Coherence") introduces the final component of question dynamics—the arising of questions—and wraps up Part III with an account of the coherence of narratives. I began by reviewing the semantic account of the arising of questions in the existing literature, according to which a question arises in an investigation just in case it is meaningful to pose in the investigation. I argued that the semantic account provides a necessary condition for questions to arise, but is insufficient to capture the sense in which questions arise by belonging to the investigative agenda. Instead, I proposed the following alternative account:

> A question Q arises at time t in an investigation, if and only if:
>
> - Q can be meaningfully posed at time t. That is, no presupposition of Q is ruled out in the investigation at time t.
> - Q possesses at least some degree of significance for the investigation at time t.

In addition, I proposed the following account of what it means for something to give rise to a question. Here X could be a declarative statement, a question, or a narrative:

> X gives rise to a question Q at time t in an investigation, if and only if:
>
> - Q can be meaningfully posed at time t. That is, no presupposition of Q is ruled out in the investigation at time t.
> - There exist some facts F about X at time t, such that F endow Q at least some degree of significance for the investigation at time t.

My account of the three main components of question dynamics—the resolution, the significance, and the arising of questions—allows me to describe the full structure of question dynamics in event reconstruction research. At each time during the investigation, the question dynamics is defined by a set of parameters, including a set $\mathcal{O}$ of open questions, a set $\mathcal{R}$ of resolved questions, and a set $\mathcal{S}$ of answers to the resolved questions so far. Moreover, each question in either $\mathcal{O}$ and $\mathcal{R}$ is associated with a degree of significance, which combines all the four dimensions of question significance; and each question in $\mathcal{R}$ is associated with a degree of the provisionality of resolution, which measures how provisional the resolution of the question is.

The question dynamics is an iterative process that repeatedly updates its parameters. At each point, new questions arise and go into $\mathcal{O}$; some old questions are resolved and go into $\mathcal{R}$, while their answers go into $\mathcal{S}$; and some previously resolved questions in $\mathcal{R}$ become open again. At the end of the investigation, the set of answers to the resolved questions in $\mathcal{S}$ is converted into a narrative N of a past event. The narrative N is based on, but not equivalent to, the set of answers $\mathcal{S}$ due to additional aesthetic and rhetorical qualities of the narrative.

Finally, I define the **coherence** of the narrative N as a measure of the overall evidential status of $\mathcal{S}$; that is, how well the answers in $\mathcal{S}$ are supported by the total evidence obtained in the investigation. Moreover, the degree of coherence of the narrative N is determined in the following way:

- The greater the size of $\mathcal{R}$ is, the higher the degree of coherence is.
- The greater the size of $\mathcal{O}$ is, the lower the degree of coherence is.
- The higher the degree of significance of any question in $\mathcal{R}$ is, the higher the degree of coherence is.
- The higher the degree of significance of any question in $\mathcal{O}$ is, the lower the degree of coherence is.
- The less provisional the resolution of each question in $\mathcal{R}$ is, the higher the degree of coherence is.
- The greater the *negative correlation* between the degree of significance of each question in $\mathcal{R}$ (on the one hand) and its degree of the provisionality of resolution (on the other hand), the higher the degree of coherence is.

This account of coherence implies that narratives that have higher degrees of coherence tend to be produced by question dynamics with specific structural features. For instance, questions with high degrees of Collingwood significance tend to be better resolved, and p-predicaments tend to be absent in more coherent narratives. I argued that paying more attention to structural characteristics of question dynamics can produce a better understanding of the evidential status of narratives, and the extent to which better-supported narratives are less vulnerable to radical revisions in the future.

17.2 A Few Comments About Generalizability

So far, I have presented my account of question dynamics and coherence in the context of event reconstruction research in general. Yet, the only examples I have examined are from engineering failure investigations. This raises the question: To what extent is my account applicable to other types of event reconstruction research that are not engineering failure investigations?

The question about generalizability is further motivated by the salient differences between engineering failure investigations and other types of event reconstruction research. For concreteness, compare the NTSB's investigations of the five aviation accidents examined so far with the research into the Cretaceous-Paleogene (K-Pg) extinction event:

First, engineering failure investigations tend to have relatively well-defined beginnings and ends, and the temporal duration of the investigations is relatively short (from a few months to a few years). This is generally not true of event reconstruction researches in other fields, which often do not have clear beginnings and ends, and can span decades or even centuries. For instance, the historical reconstruction of the K-Pg extinction seeks to identify the causes of an unusually drastic turnover in the preserved fossil types in a thin layer of sediment called the K-Pg (or K-T) boundary. It is not straightforward to determine precisely when this research program began, although it presumably dated back to the early decades of the 20th century. It is still an active research program with recent discoveries.

Second, engineering failure investigations tend to follow relatively well-defined processes and procedures; consequently, the structures of the question dynamics in these investigations tend to be routine and are relatively straightforward to identify. For instance, after securing an aircraft's wreckage, NTSB investigators would divide into groups and conduct a comprehensive search for anomalies in all aspects of the accident flight. After identifying all (or nearly all) the anomalous traces, the investigators would raise questions about each anomalous trace to determine whether it contributed to the accident. Moreover, the availability of the black boxes, radar return records, and air traffic communication records allowed the investigators to construct a partial chronology of the event early in the investigation. Detailed examinations of the secured wreckage enabled them to fill in further details of the chronology. Having a fairly detailed chronology, in turn, helped the investigators to focus on questions about how the failure sequence initiated and what had contributed to the initiation.

In contrast, event reconstruction research in many other fields tends to have more complex discovery processes and less routine question dynamics. This is due to various factors. First of all, it is generally impossible to collect all the traces at the beginning of these investigations. In fact, the investigators may not even know *what* traces to look for and *where* to look for them at a given point in the investigation. For instance, the need to look for a suitably dated giant impact crater only became salient when the Alvarez's impact hypothesis was taken seriously. Even then, it was not clear where to look for such a crater, and it took a decade before Hildebrand and his colleagues announced the discovery of the Chicxulub Crater at the K-Pg boundary. ([21], p.4)

Furthermore, the question dynamics of many event reconstruction researches were often shaped by contingent historical factors, such as the emergence of new technologies that enabled access to certain new traces, or accidental discoveries made in other seemingly unrelated researches. For instance, the discovery of the 30-fold enrichment of iridium in the K-Pg boundary layer—a major piece of evidence supporting the impact hypothesis—was accidental: The Alvarez team initially set out to measure the iridium concentrations in the K-Pg boundary to determine the length of time represented by this sediment layer; the measurement, in turn, was made possible by the recently observed correlation between sedimentation rate and iridium concentration. ([21], p.4)

Third, engineering failure investigations are fundamentally practically oriented, with very clearly defined practical goals. Improvements in the safety of the engineering systems and prevention of similar future accidents are arguably the most important goals of any engineering failure investigation. Consequently, many of the questions that arise in engineering failure investigations derive their significance from their relevance to fulfilling some of the practical goals of the investigations. The overall structures of question dynamics are also heavily influenced by these practical goals. For instance, an engineering failure investigation typically terminates when the investigators have identified a sufficient number of contributing factors to the accident that could prevent similar accidents from occurring again.

In contrast, the vast majority of event reconstruction research in other fields is much more intellectually oriented, with little to no practical benefits. For instance, it is unclear what the practical payoffs of a reconstruction of the K-Pg extinction event are. Consequently, the significance of questions that arise in these researches and some of the structural characteristics of the question dynamics, should be different from those in engineering failure investigations.

Fourth, event reconstruction research in many other fields tends to be characterized by higher levels of disagreements than engineering failure investigations. This is not to deny that disagreement can occur in engineering failure investigations: For instance, individual investigators can (and do) disagree with each other during an investigation, and we have seen disagreements among investigative agencies, such as the disagreement between the NTSB and the FBI in the initial investigation of TWA 800 accident, and the disagreement between the NTSB and Boeing in the investigation of USAir 427 accident. However, at least for high profile investigations such as plane crash investigations, disagreements tend to dissolve over time as evidence accumulates, and significant disagreements that can result in unresolved cases are rare. Consequently, it is generally feasible to rationally reconstruct the investigative processes without explicitly representing the disagreements.

In contrast, disagreements in many other types of event reconstruction research tend to be much more persistent. For instance, so far, the *existence* of an impact event by an asteroid has been very well established by researchers of the K-Pg mass extinction event. However, the precise causal relationships between this impact event and the mass extinction of the biological species are still controversial. Some researchers argued that the impact event was sufficient to account for the scope and selectivity of the extinctions; other researchers argued that massive geological processes such as volcanism and marine regression also contributed to the extinction. Even the timing of the impact event relative to the time scale of the mass extinctions has not yet been determined with enough precision for settling the causal questions.

Event reconstruction researches in fields such as history and social sciences arguably have even higher levels of disagreements. In some cases, the disagreements are sufficiently entrenched that researches on a given topic divide into a variety of competing research paradigms or schools, with researchers in different paradigms asking different but related questions about the event to be reconstructed. In these cases, it is unclear when or whether the disagreements will ever be resolved in the future. Consequently, it seems that disagreements are indispensable parts of the question dynamics in these types of event reconstruction research.

It follows that there are important differences between the question dynamics of engineering failure investigations and the question dynamics of other types of event reconstruction research, and my account of the former does not fully generalize to the latter. I do not dispute this conclusion. However, this conclusion does not imply that my focus on engineering failure investigations so far is misguided. In fact, the *point* of using engineering failure investigations as paradigm cases of event reconstruction research is *precisely* that they are in many ways simpler, and thus easier to study and characterize, than other types of event reconstruction research. By focusing on question dynamics with simpler structures, we can develop some useful abstractions and conceptual resources, which can be modified and expanded when we examine more complex question dynamics of other types of event reconstruction research.

For instance, since my notion of Collingwood significance of a question is defined at the end of the investigation, it does not apply to other types of event reconstruction research that do not have clear endings. However, it is straightforward to define variants of this notion relative not to the end of the investigation, but to any particular moment during the investigation. Similarly, perhaps my distinction between the two types of evidential questions, or the distinction among the four dimensions of question significance, are not sophisticated enough to capture the full complexity of evidential questions or question significance in other types of event reconstruction research. But the general *framework* of distinguishing between different forms of evidential questions and distinguishing between different dimensions of question significance is likely generalizable.

Finally, even complications such as persistent disagreements in certain event reconstruction researches may be characterizable within the general framework of question dynamics. We could introduce additional conceptual apparatus to represent the questions that arise in each competing research paradigm, and how the questions in different research paradigms are connected. Exactly what conceptual apparatus should be introduced, of course, will depend on detailed examinations of the histories and practices of particular event reconstruction research programs.

Bibliography

[1] Bill Adair. *The Mystery of Flight 427: Inside a Crash Investigation*. Smithsonian, 2004.

[2] National Aeronautics and Space Administration. “Deadly Efficiency”. In: *System Failure Case Studies* 4.7 (2010), pp. 1–4.

[3] National Aeronautics and Space Administration. “Fire in the Sky”. In: *System Failure Case Studies* 5.1 (2011), pp. 1–4.

[4] National Aeronautics and Space Administration. “No Left Turns”. In: *System Failure Case Studies* 2.6 (2008), pp. 1–4.

[5] US Airways. *USAir 427: US Airways’ View of the Accident*. URL: `https://www.avweb.com/features/usair-427-us-airways-view-of-the-accident/`.

[6] George Bibel. *Beyond the Black Box: the Forensics of Airplane Crashes*. JHU Press, 2008.

[7] George Bibel and Captain Robert Hedges. *Plane Crash: The Forensics of Aviation Disasters*. JHU Press, 2018.

[8] National Transportation Safety Board. *Aircraft Accident Report: American Airlines, INC. DC-10-10, N110AA, Chicago-O’Hare International Airport, Chicago, Illinois (May 25, 1979)*. Tech. rep. NTSB-AAR-79-17. Washington, DC, 1979.

[9] National Transportation Safety Board. *Aircraft Accident Report: Uncontrolled Descent and Collision With Terrain, United Airlines Flight 585, Boeing 737-291, N999UA, 4 Miles South of Colorado Springs Municipal Airport, Colorado Springs, Colorado (March 3 1991).* Tech. rep. NTSB/AAR-01/01. Washington, DC, 2001.

[10] National Transportation Safety Board. *Aircraft Accident Report: Uncontrolled Descent and Collision With Terrain, USAir Flight 427, Boeing 737-300, N513AU, Near Aliquiipa, Pennsylvania (September 8 1994).* Tech. rep. NTSB/AAR-99/01. Washington, DC, 1999.

[11] National Transportation Safety Board. *Aircraft Accident Report: United Airlines Flight 232, McDonnell Douglas DC-10-10, Sioux Gateway Airport, Sioux City, Iowa (July 19, 1989).* Tech. rep. NTSB/AAR-90/06. Washington, DC, 1989.

[12] National Transportation Safety Board. *Aircraft Accident Report: United Airlines Flight 585, Boeing 737-291, N999UA, Uncontrolled Collision With Terrain For Undetermined Reasons, 4 Miles South of Colorado Springs Municipal Airport, Colorado Springs, Colorado (March 3 1991).* Tech. rep. NTSB/AAR-92/06. Washington, DC, 1992.

[13] National Transportation Safety Board. *Aircraft Accident Safety Report: In-flight Breakup Over the Atlantic Ocean, Trans World Airlines Flight 800, Boeing 747-131, N93119, Near East Moriches, New York (July 17, 1996).* Tech. rep. NTSB/AAR-99/01. Washington, DC, 2000.

[14] National Transportation Safety Board. *NTSB Board Meeting on TWA 800, Day 1. August 22, 2000.* URL: `https://en.wikisource.org/wiki/Portal:TWA_Flight_800_investigation/Day1`.

[15] National Transportation Safety Board. *NTSB Board Meeting on TWA 800, Day 2, Part 3. August 22, 2000.* URL: `https://en.wikisource.org/wiki/Portal:TWA_Flight_800_investigation/Day2-3`.

[16] National Transportation Safety Board. *The Investigative Process.* URL: `https://www.ntsb.gov/investigations/process/Pages/default.aspx`.

[17] Sylvain Bromberger. *On What We Know We Don't Know: Explanation, Theory, Linguistics, and How Questions Shape Them.* University of Chicago Press, 1992.

[18] Gerry Byrne. *Flight 427: Anatomy of an Air Disaster.* Springer Science & Business Media, 2002.

[19] Adrian Currie. *Rock, Bone, and Ruin: An Optimist's Guide to the Historical Sciences.* MIT Press, 2018.

[20] Key Dismukes, Benjamin A Berman, and Loukia D Loukopoulos. *The Limits of Expertise: Rethinking Pilot Error and the Causes of Airline Accidents.* Ashgate Publishing, Ltd., 2007.

[21] Patrick Forber and Eric Griffith. "Historical Reconstruction: Gaining Epistemic Access to the Deep Past". In: *Philosophy and Theory in Biology* 3.201306 (2011), pp. 1–19.

[22] Peter Galison. "An Accident of History". In: *Atmospheric flight in the twentieth century.* Springer, 2000, pp. 3–43.

[23] Andrew Gelman and Guido Imbens. *Why Ask Why? Forward Causal Inference and Reverse Causal Questions.* Tech. rep. National Bureau of Economic Research, 2013.

[24] Captain AI Haynes. *Eyewitness Report: United Flight 232.* URL: `https://web.archive.org/web/20080618091811/http://www.airdisaster.com/eyewitness/ua232.shtml`.

[25] Captain AI Haynes. *The Crash of United 232.* URL: `http://clear-prop.org/aviation/haynes.html`.

[26] James Lamiell. "Nomothetic and Idiographic: Contrasting Windelband's Understanding with Contemporary Usage". In: *Theory & Psychology* 8.1 (1998), pp. 23–38.

[27] John Mackie. “Causes and Conditions”. In: *American Philosophical Quarterly* 2.4 (1965), pp. 245–264.

[28] Christine Negroni. *Deadly Departure: Why the Experts Failed to Prevent the TWA Flight 800 Disaster and How It Could Happen Again*. Harper Collins, 2013.

[29] L.A Paul. “Aspect Causation”. In: *The Journal of philosophy* 97.4 (2004), pp. 235–256.

[30] Nicholas Rescher. *Inquiry Dynamics*. Transaction Publishers, 2000.

[31] Alexander Reutlinger, Gerhard Schurz, and Andreas Hüttemann. “Ceteris Paribus Laws”. In: *Stanford Encyclopedia of Philosophy* (2011).

[32] William Robbins. *From Jet disaster, an Iowa Family Harvests a Reward*. URL: `https://www.nytimes.com/1989/10/14/us/from-jet-disaster-an-iowa-family-harvests-a-reward.html`.

[33] Neville Sachs et al. *Practical Plant Failure Analysis: A Guide to Understanding Machinery Deterioration and Improving Equipment Reliability*. CRC Press, 2019.

[34] Ed Stiles. *Engineering Alum Helped Crack Mystery of Flight 427*. URL: `https://uanews.arizona.edu/story/engineering-alum-helped-crack-mystery-flight-427`.

[35] The Seattle Times. *Safety At Issue: The 737, Part 2*. URL: `https://special.seattletimes.com/o/news/local/737/part02/index.html`.

[36] The Seattle Times. *Safety At Issue: The 737, Part 3*. URL: `https://special.seattletimes.com/o/news/local/737/part03/index.html`.

[37] Bas Van Fraassen. *The Scientific Image*. Oxford University Press, 1980.

[38] Mara Vatz. “Knowing When to Stop: the Investigation of Flight 191”. MA thesis. Massachusetts Institute of Technology, 2004.

[39] Andrzej Wisniewski. *The Posing of Questions: Logical Foundations of Erotetic Inferences*. Vol. 252. Springer Science & Business Media, 2013.

[40] James Woodward. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, 2005.