The Challenger Disaster

Daniel Hastings
September 2003

Overview

The explosion of the Space Shuttle Challenger in 1986 has a specific proximate cause that can be traced to a set of interlocking technology and policy decisions over many years going back to the late sixties.

Technology Issues

Space technology: Large payload needs (40, 000 lbs up-mass) to orbit and the rocket equation (follows from conservation of mass and momentum for a rocket) and the deep gravity well on the surface of the Earth demand very high thrust rockets for takeoff (millions of lbsf). Economic development considerations drove the choice of segmented solid rocket boosters to give the initial high thrust. Solid rockets burn solid fuel and oxydiser at very high temperatures and cannot be turned off once lit. Thus once a leak developed in an O-ring (necessary because it was segmented) the fate of the Challenger and the astronauts was sealed.

Policy Issues

Space policy: The US Space Policy statement of 1982 defined the Shuttle (STS) as the US primary means of space transportation and said that it would be both fully operational and cost effective in providing routine access to space. Fully operational meant that it could be used for routine operations (not test flights) and cost effective was interpreted to mean that it could launch on schedule without delays.

The policy statements put pressure on the NASA managers to use a technically unsound design and to launch in extreme conditions. This was the backdrop to the actual specific cause of the disaster. This illustrates one of the issues with complex systems, that some behavior may be emergent since the sequence of events that led to the disaster was not predicted by anyone.

Reading

Launius , Roger D. and McCurdy, Howard E. *Spaceflight and the Myth of Presidential Leadership*. Urbana: University of Illinois Press, 1997. (Especially Chaps 2-6)

Introduction to Systems Thinking (10pp) [Sterman, John *Business Dynamics: Systems Thinking and Modeling for a Complex World,* Irwin/McGraw-Hill, 2000 and other information is generally available at http://www.hps-inc.com]

Judgment Under Uncertainty (25 pp) [Kahneman, D, Slovic, P. and Tversky, A. (1982) Judgment under Uncertainty: Heuristics and Biases. New York: Cambridge University Press, pp. 463-489.]

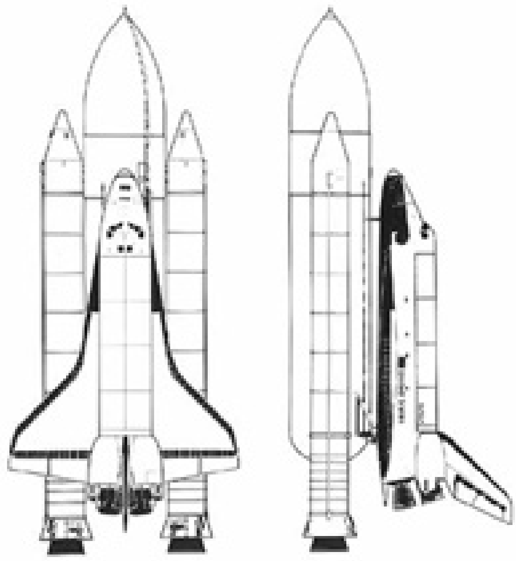
An Outsider's Inside View of the Challenger Inquiry (11pp) [Feynman, Richard (1988) "An Outsider's Inside View of the Challenger Inquiry," *Physics Today*, February, pp. 26-37.]

Themes

The proximate cause of the Challenger disaster

On 28 January, 1986 the Space Shuttle Challenger took off with a teacher on board and exploded 73 seconds later. The immediate cause of the explosion was a burn through of one of the O-rings on one of the solid rocket boosters. This caused the solid rocket (steel) wall to fail at the burn-through point. The solid rocket then pivoted into the large external tank causing release of hydrogen which underwent a deflagration leading to the Shuttle Challenger being ripped apart at altitude.

The proximate cause was the leakage of two rubber O rings in a segmented solid rocket booster. The rings has lost their ability to stop hot gas blow-by because on the day of launch they were cold (estimated at 20 degrees F, well below freezing). The ambient temperature at launch was in the low 30s. See Figs 1-5 which show the location of the boosters on the Shuttle, the segmented design of the solid rocket boosters, the detailed design and location of the O-rings, the start and end of the explosion.



Courtesy of NASA.

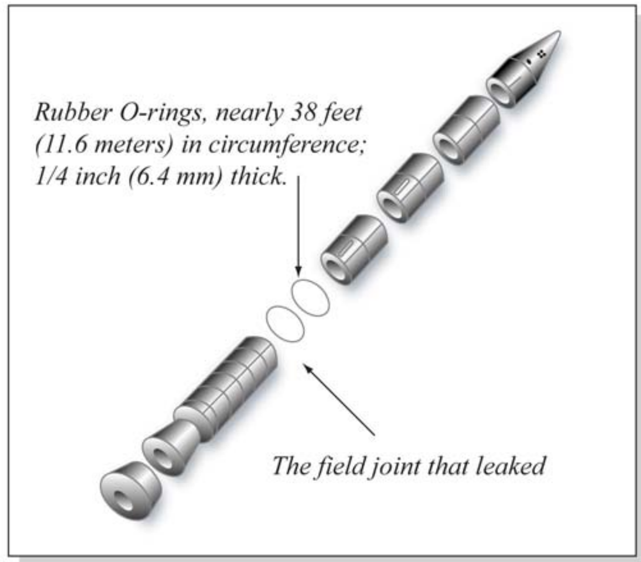


Figure by MIT OCW.

Figure 1. Space Shuttle and Solid Rocket Booster Design showing location of O-Rings

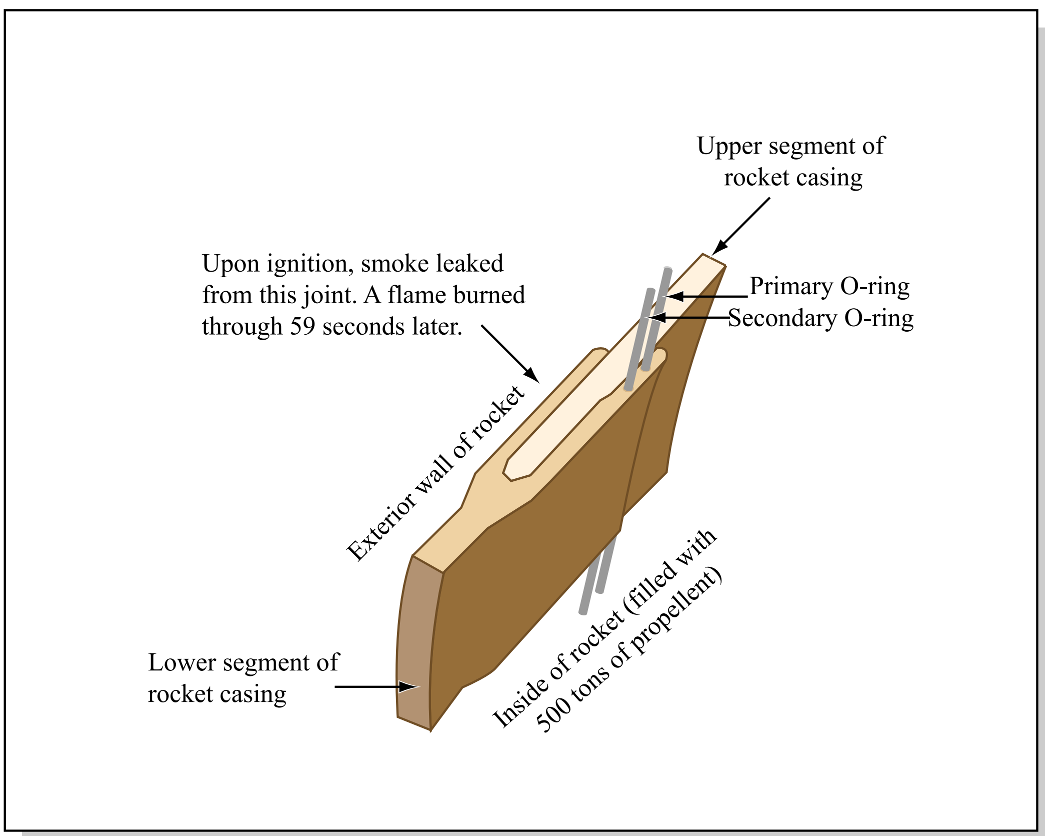


Figure by MIT OCW.

Figure 2. Detailed Location of O-Rings

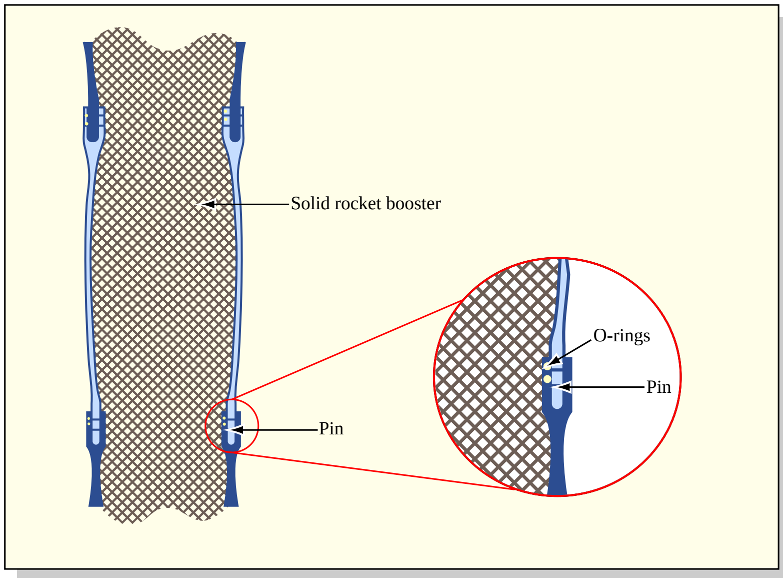
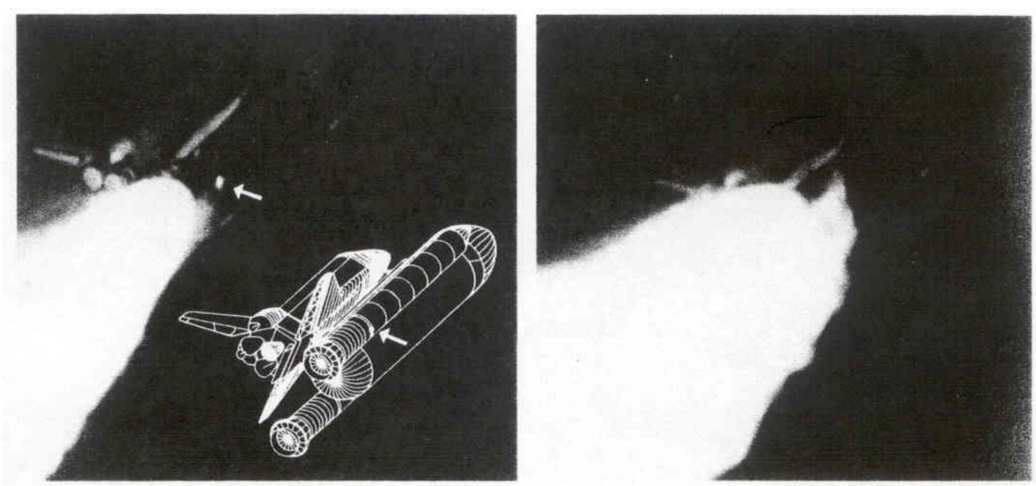


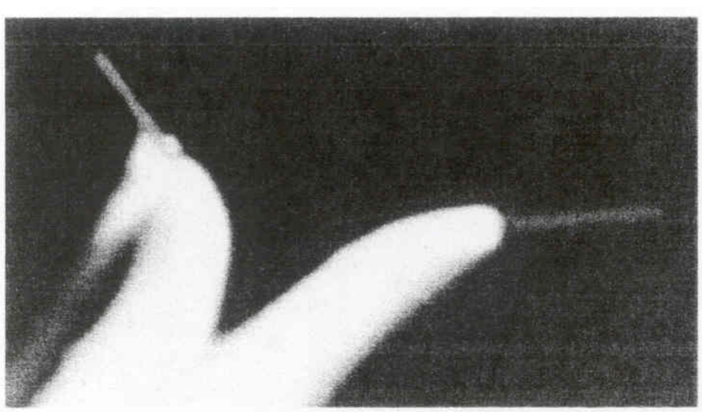
Figure by MIT OCW.

Figure 3. Detailed Design of O-Rings in Solid Rocket Booster



Courtesy of NASA.

Figure 4. Picture showing the start of the explosion



Courtesy of NASA.

Figure 5. Picture showing the end of the explosion - The Shuttle is gone

Amazingly the exact cause of the accident was debated for hours the evening before the launch between Morton Thiokol engineers, managers and NASA managers. Given the predicted temperatures of 26 degrees F, the engineers were concerned that the O rings might not be resilient and that there was a history of O ring erosion on the STS during cold weather launches. This led them to recommend that the STS not launch at these low temperatures. This was the first no-launch recommendation from Morton Thiokol in the history of the STS. Initially, the Thiokol managers supported the engineers. But under disbelieving questioning by the NASA managers, the Thiokol managers put on their management hats (their phrase!), changed their minds and changed the Thiokol recommendation to launch. The NASA managers were thus mollified and felt justified in approving a launch with the well-known result that Challenger exploded.

While the proximate cause was debated at length the night before, the disaster was also the result of a set of coupled, feedback loops between technical, economic and policy decisions that made such an accident almost inevitable. These decisions set the stage for the specific cause. In the case we will address those coupled decisions as well as the communication and ethics issues associated with the specific cause. These decisions illuminate the following questions.

Technical

1. Why were solid rockets (solid high explosive) chosen to launch people when the known failure rate of solid rocket boosters was 1 in 25 (based on years of launching solid rockets)?

2. Why were the solid rocket boosters designed with segments thus introducing potential leak paths and necessitating O-rings?

3. How was the failure rate of the Shuttle assessed at 1 in 10,000 when key components (the solid rockets) had failure rates of 1 in 25?

Economic

1. Why was the Shuttle development cost capped even though it was known it would lead to higher long term operating costs?

2. Why was the solid rocket booster manufactured in a site that could not ship a whole solid rocket booster?

Policy

1. Why was the Shuttle made the primary US launch vehicle by policy?
2. Why was it declared an operational vehicle after only four test flights?
3. Why was the policy decision made to allow a teacher on board the flight?

Poor Communication and Poor Ethics

In the investigation that followed a number of contributing factors were identified. First, NASA managers, under pressure to show the STS was reliable, had authorized a launch even though the temperature criteria were outside of the known operational range of the STS. In a sense the operational mindset, dictated by policy, had over taken them. They over ruled the engineers who warned of possible danger. Second, NASA and Morton Thiokol engineers had known for some time that there were problems with gas blow-by through the O-rings. However, the NASA system ignored these signs and did not calculate the consequences of a blow-by in a way that was clearly communicated to senior management. The next figure shows the actual data that was presented to the Morton Thiokol and NASA senior managers at the discussion the night before.

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

			Cross Sectional View			Top View		
		SRM No.	Erosion Depth (in.)	Perimeter Affected (deg)	Nominal Dia. (in.)	Length Of Max Erosion (in.)	Total Heat Affected Length (in.)	Clocking Location (deg)
Oct 30, 1985	61A LH Center Field** (AFT)	22A	None	None	0.280	None	None	36°--66°
	61A LH ~~CENTER~~ FIELD**	22A	NONE	NONE	0.280	NONE	NONE	338°-18°
85	51C LH Forward Field**	15A	0.010	154.0	0.280	4.25	5.25	163
	51C RH Center Field (prim)***	15B	0.038	130.0	0.280	12.50	58.75	354
	51C RH Center Field (sec)***	15B	None	45.0	0.280	None	29.50	354
	41D RH Forward Field	13B	0.028	110.0	0.280	3.00	None	275
	41C LH Aft Field*	11A	None	None	0.280	None	None	--
	41B LH Forward Field	10A	0.040	217.0	0.280	3.00	14.50	351
July	STS-2 RH Aft Field	2B	0.053	116.0	0.280	--	--	90

*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port - 0 deg.

OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.

SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO SOOT BLOWBY. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY.

Courtesy of NASA.

Figure 6. Actual presentation of the O-ring damage data before the Challenger launch

This presentation of the data shows that on some STS flights there had been damage to the primary and secondary O-rings. Note this that this alone should have given pause for concern since the technical design called for no damage to the primary and absolutely no damage to the secondary. However, it does not show any temperature dependence (the key variable) or give any indication of how serious the erosion actually was. Thus the engineers, who were very concerned since they knew the consequences of blow-by of the secondary ring, did not present the data in a way that showed the trends or that could be explained to a skeptical senior NASA manager. They thought they were communicating but missed one of the fundamental rules of good communication which is to express your position in a way that your customer can understand. The next figure shows how the data could have communicated (note they had all of this data on hand in the discussion).

Figure removed due to copyright restrictions.
Plot of O-ring damage index vs. Temperature at time of launch.
Source: Tufte, Edward. V*isual Explanations: Images and Quantities, Evidence and Narrative.* Cheshire, CT: Graphics Press, February 1997.
ISBN: 0961392126.

Figure 7. How the O-ring damage data could have been presented before the Challenger launch

This presentation of the data makes anumber of things clear. It suggests that there is damage to the O-rings associated with several flights. It also suggests that a reasonable correlation is that the damage increases with decreasing temperature although there is some uncertainty associated with the data. However, it does show that <u>every</u> launch below 66F resulted in damaged O-rings. Furthermore it shows that the predicted temperatures for the launch were well outside the previous experience base that NASA had with this complex machine. This would have communicated to senior decision makers much better the dangers and could have been shown to the NASA administrator who would have made the final launch/no-launch decision if he had been at the Cape.

However, the NASA communication system by this time was so poor that senior managers did not know of these potential issues and the NASA administrator for the first time ever did not go to the Cape for the launch. He thought this was a routine launch of an

These factors point up issues of communication and ethics. Even though there was great danger, no one in the system felt empowered to listen and act. The managers ignored the experts and did not allow multiple ways of checking on these critical systems. There should have been a communication system whereby the engineers could have spoken to the NASA managers and caused an independent review of the relevant data (on the grounds that two independent sets of eyes are better than one). In addition, the engineers should have been willing to resign over an issue where the stakes were so high. Every engineer and decision maker needs to understand what is his or her bottom line with respect to decisions. When the bottom line is crossed, then the ethical choice is to separate oneself from the decisions. This is fundamentally a question of values based on integrity & technical excellence. When a critical decision is imminent is too late to decide on what values are important. The willingness to separate oneself from flawed decisions also clearly demarcates the boundaries for the decision maker.

Dynamic Complexity

The STS Challenger accident is an example of dynamic complexity making a prediction of the behavior of a system very hard. Also called behavioral complexity, this is basically the degree to which the outputs (behavior) of a system are difficult to connect to the inputs.[1[1]] In this case, decisions made in the late sixties helped set the stage for an accident fifteen years later.

The definition may not sound striking, but the concept is critical. Intuitive understanding, formal modeling, and rational decision-making are all fundamentally based upon our expectations about how system behavior will change in response to specific changes in inputs. If our beliefs about these relationships are incorrect, we cannot use available information to make wise decisions. This is actually fairly common: there are numerous examples of the process, sometimes called "policy resistance," of believing we understand a problem, taking steps to try to remedy it, but failing (and too often making things worse): anti-lock brakes in cars make drivers more aggressive, decreasing safety; fire-suppression policy results in more severe forest fires; low-tar cigarettes lead smokers to smoke more, increasing carcinogen intake, etc. Except in very special cases, however, there is not really an aspect of the system deliberately resisting the intended solution. Rather, our understanding of the system – specifically, the reasons for its behavior – is inadequate. Why? Many factors contribute to dynamic complexity. Six of the most important (commonly present but unrecognized or misunderstood) are introduced below.

[1[1]] For the concept of dynamic complexity and some characteristics of dynamically complex systems, we are indebted to John Sterman. For more information, see Business Dynamics: Systems Thinking and Modeling for a Complex World (Irwin/McGraw-Hill, 2000).

Dynamics – over a long enough time frame, almost any aspect of a system can change; in a short enough period, none do. Unfortunately, key elements can change at dramatically different rates, making them difficult to judge. Causes and effects may be distant in time, and their relationships may change by the time their connections are recognized.

Coupling – components of a system are said to be "coupled" if changing one affects the other. Tightly coupled elements make it almost impossible to affect the intended target exclusively; any change also causes some other changes, each of which may cause still more changes, … with each induced effect being less intentional and therefore potentially less predictable, recognizable, and controllable.

Feedback – the condition created by a causal loop of couplings. When a system involves feedback, a known change to a factor can eventually travel back and exert an additional change to the original factor. A single feedback loop can result in significantly greater or lesser effects than expected, if the feedback is not anticipated. Multiple loops, acting in different time frames, can result in practically unpredictable behavior.

Nonlinearity – when the changes in one or more outputs are not proportional to changes in the inputs. This is caused by coupling or feedback, and is primarily problematic because people rarely anticipate – or fully understand – nonlinear relationships.

Chaos – for our purposes, this is basically unpredictable behavior from a deterministic system. Chaotic systems are characterized by extreme sensitivity to initial conditions and unpredictable, aperiodic evolution (with some stable structure) instead of convergence to a steady-state. The classic description of chaotic behavior is the "butterfly effect."

Adaptation – if the capabilities or preferences of actors within the system can evolve, (i.e., they learn from or actively respond to events), self-organization, self-selection, and co-evolution can result in sophisticated "emergent" behavior.

Failure to recognize these characteristics in a system may lead to underestimation of the dynamic complexity of the system. Consequently, predictions of system behavior may be seriously flawed, while unwarranted confidence in these predictions (stemming from the belief that the "mechanics" of the systems are understood) may be maintained.

<u>The Technical Design of the STS & the path to a segmented solid rocket design</u>

When President Nixon took office in 1969, NASA funding was already going down (see Figure 8).

The Technical Design of the STS & the path to asegmented solid rocket design

When President Nixon took office in 1969, NASA funding was already going down (see Figure 8).

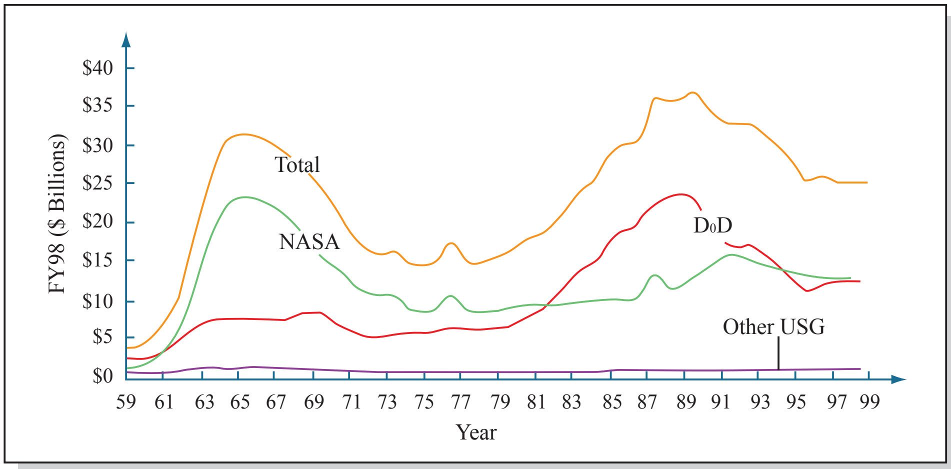


Figure by MIT OCW.

Figure 8: NASA Funding

The first Moon landing occurred in July 1969. The race to the Moon with the Soviets was won! It was like the dog that caught the truck. What would NASA do now? To some extent NASA was caught in a time warp. NASA felt that after the first lunar landing it should get whatever funding it needed. In September 1969, a Space Task Group chaired by Vice President Agnew reported three possible long-range space programs for NASA. The first was a manned mission to Mars by mid-eighties, an orbiting lunar station and a fifty man Earth orbiting station served by a reusable shuttle. Funding for this option was $8 to $10 billion/yr. (Recall that at its peak NASA had received 5 billion/yr, see figure above, thus this was an a doubling of the annual NASA budget). The second plan postponed Mars until 1986 and limited funding to $8 billion/yr. The third plan chose only the space station and shuttle, with annual spending between $4 billion-5.7 billion/yr. However relative to the long gone days of the early sixties, the mood of the country and of the President had changed. Nixon came from the Eisenhower mentality that saw the big manned effort as stunts and saw the national security space program (strategic reconnaissance satellites) as more important than the NASA effort. He was also program (strategic reconnaissance satellites) as more important than the NASA effort. He was also much more interested in promoting public cooperation rather than competition with the Soviets and the Chinese. Further he strongly believed in frugality in government spending. All these combined to make him cast a skeptical eye on the NASA requests.

The country also had changed. In 1969, the US had reached the Moon. The national mood was to turn to other issues especially in light of riots in cities, the war in Vietnam, etc. Flights to the Moon seemed boring. For NASA it was a boom or bust cycle. As a measure of this, the Congress reorganized the standing space committees out of existence and Nixon abolished the Presidential Science Advisory Committee that had been instrumental in advocating the Apollo program. Space became a secondary issue for the political establishment. Thus the last two Apollo flights were cancelled, the Apollo Application Program was reduced to one SKYLAB and in a blow to the Air Force the Manned Orbiting Laboratory (the first US Space Station) was cancelled. President Nixon refused to support any of the options that NASA wanted. There was no congressional support for any big new initiative so NASA started to wither.

It was only the 1972 election that saved something for NASA. The declining population in the aerospace industry in the big states of California, Texas and Florida forced the President to approve something for NASA so that he could blunt the criticism of the Democrats in these big states. He chose half of half of option 3. The choice was for a Space Transportation System (STS), a space truck but the place it was to go to was cancelled. That is, the Space Station was not approved. Thus it was a space truck to nowhere (this led to many conflicting requirements). It was even worse than that. NASA had suggested a completely reusable design based around liquid rocket engines (which can be actively throttled and turned off). The idea was to stop throwing away expensive hardware. Nixon would only give them half the money requested. Thus they did away with the completely reusable design and even worse with the liquid rocket engines. In a compromise to fit within a fixed $3.2 billion NASA budget, they chose a non-reusable main tank and worst of all, to make up the thrust they chose solid rocket motors.

As an aside, Von Braun had said that no human should ever ride on solid rockets. They were just too dangerous. One in twenty-five blew up due to defects. They could not be stopped once lighted and thus had the potential for a major loss of life. However, to reduce development costs of liquid rocket engines, NASA chose to go with solid rockets. In another first, they chose to go with Morton Thiokol, from the home state of the NASA administrator. Morton Thiokol was in Utah, which is where it manufactured the solid rocket segments. Of course, by this choice they also brought another state into the fold to support the President. However a completed solid rocket would be too big to transport by road to a port on the Gulf of Mexico to get it over to Cape Canaveral in Florida. Thus it had to be built in segments (which needed to be sealed with O-rimgs) and integrated at Cape Canaveral. Thus the seeds were sown for the Challenger disaster of a decade or so away. As a continuation of the sixties mindset of higher, faster and farther, NASA chose to develop shuttle main engines which had the highest thrust to weight ratio of any ever built. They would be wonders of technology. It was argued that each engine would be reusable for 100 flights and that the shuttle would fly 100 times a year. In the operational phase the cost for launch was supposed to be only \$10 million a flight. Since its payload was 40000 lbs. to Low Earth Orbit (LEO) it would give cost of \$250/lb to LEO.

However even then some issues were seen. Since the STS could only go to LEO (~250km in altitude) it would have to carry an upper stage for it to be useful for any other obit. NASA thus sold itself to other organizations to get the support it needed. The Shuttle payload bay was sized for various military missions as well as the payload carrying capacity to LEO. It persuaded the Air Force to develop a solid propellant upper stage (IUS) to put 500 lbs. into LEO. It persuaded McDonnell Douglas to build two upper stages in return for a monopoly position. These were the PAM-D and PAM-A upper stages. It also started a cryogenic upper stage based on Centaur technology. NASA was in the desperate position (as it saw it) of having to do a big project to keep itself going and it was selling itself to get approval for the big project. The cost projections which finally sold the administration were based on a large number of flights a year which was based on a market which did not yet exist- (even today ~50 flights /yr worldwide for all types of launches). Thus there was a classic chicken-and-egg problem. In retrospect the fundamental problem was forcing a pioneering technical program to be justified in economic terms. In this sense there was a huge disconnect between NASA and the administration. Note that Apollo was never justified on economic terms.

Thus for fundamentally economic and political reasons, NASA chose a unsound technical design namely segmented solid rocket boosters to launch human beings. Note that even today Ariane 5, which launches only satellites, makes the solid rocket boosters it uses in-situ in Guyana and makes them as unitary systems without segments. They decided that introducing leak paths into solid high explosive was too risky.

Probability Basics

The most widely used formalism for classifying uncertainty is probability. The classical view is that the probability of an event occurring in a set of trials is the frequency of event occurrence. From this equation, it is clear that probability must always be a quantity between 0 (no probability of occurrence) and 1 (certain occurrence). Consider two events, A and B, with probabilities of occurrence of $P(A)$ and $P(B)$, respectively. If A and B are *independent* events – if one occurring or not has absolutely no bearing on whether the other occurs – then we know the following:

P (A <u>and</u> B) = P(A)P(B)   {a.k.a. the *intersection* of A and B}

Thus given the probability of one of the Space Shuttle Main Engines failing is 1 in 100, the probability of two engines failing (which would have led to loss of the Shuttle since one engine out is insufficient to cause Shuttle failure & assuming the failures are not due to the same cause) is 1 in 10,000 which is what they estimated. Since they were planning on 100 flights a year then the Shuttle would have flown for 100 years before there would have been a catastrophic launch. If the probability of a solid rocket failing is 1 in 25 then the probability of both solid rockets failing is 1 in 625 which would certainly cause Shuttle loss. For reference, the current NASA estimate of catastrophic Shuttle loss is 1 in 250.

<u>Risk and cost estimates for the Shuttle</u>

In contrast to the rosy market projections, the facts are that NASA has never managed more than nine STS flights a year (see Fig 4), the Shuttle Main Engines (SME) needed to be replaced every flight and the cost estimates per launch range from \$80 million to \$500 million. There are three ways to estimate cost. The first is to take the total amount spent so far on STS and divide by the number of flights. This gives about \$500 million/yr. The second is to take the annual amount in the NASA budget and divide by the annual flight rate. This gives about \$250 million/yr. The last is to ask how much is saved when an STS flight is cancelled. This is about \$80 million/yr. This last figure is telling since what are saved are only the consumables. Most of the cost is in the standing army necessary to operate and maintain the shuttle. This cost and the low reliability of the shuttle were not appreciated in the initial estimates. There was also some specious thinking at NASA about markets and either wishful thinking or an under-appreciation of the difficulty of developing a new engine. The new engine contributed to the delays of the first STS launch until 1981 and have contributed greatly to the poor reliability of the STS. A truck it is not, it is much more like a finely tuned racecar.

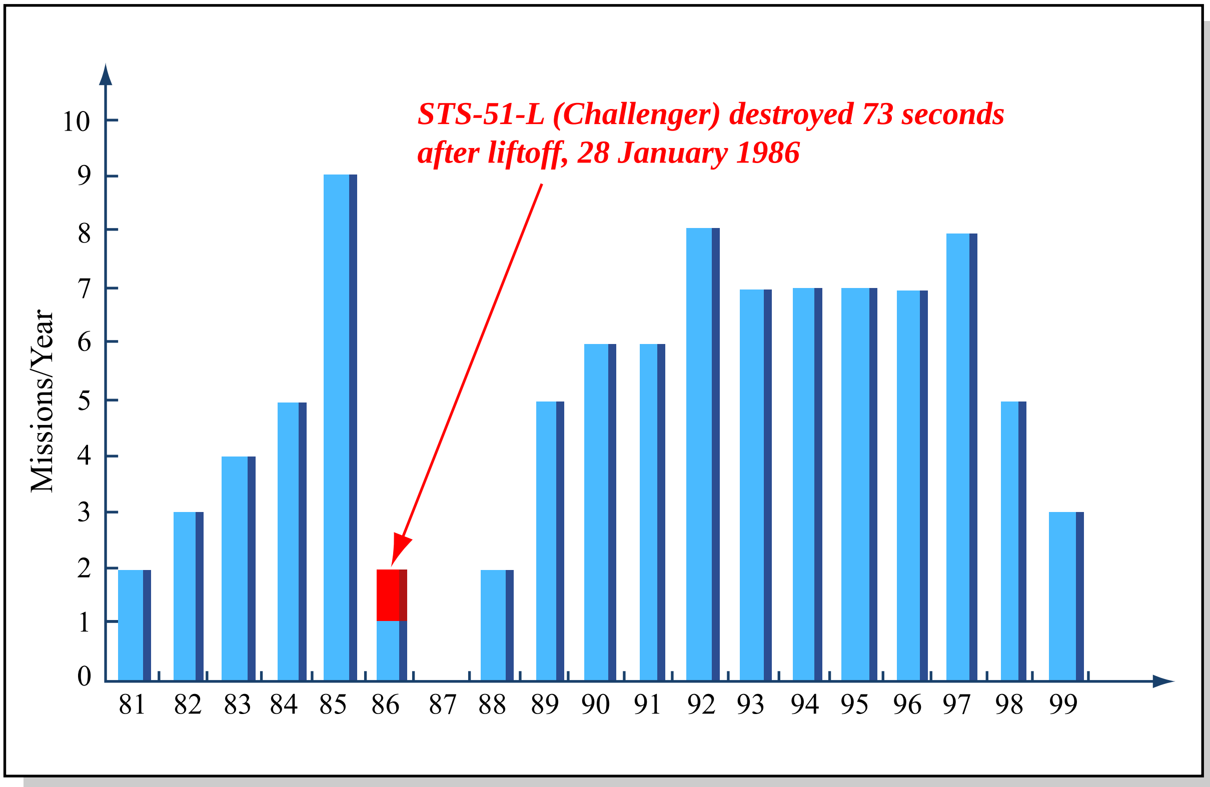


Figure by MIT OCW.

Figure 9: Number of Shuttle Launches against year

In 1977, NASA projected that the shuttle would fly 600 times in the first eleven years of operation. The failure rate was estimated at 1 in10,000 flights and the reliability (i.e. ability to take off on time) was estimated at 98%. The total cost of developing the shuttle in 1972 was estimated at $8 billion with each new orbiter costing $250 million to build. The first test flight was scheduled for early 1978. The Shuttle was designed to DOD requirements to place reconnaissance satellites in orbit and retrieve them. Thus both it's size and cross range flowed from the intelligence requirements. The facts were very different. The first shuttle flew on April 12,1981, three years late mainly due to the technical requirements and difficulties associated with the Space Shuttle main engine. It cost $12.6 billion to develop and each orbiter cost almost a billion to produce. The cost per payload pound is over $10,000. In the years 1983-1994, it only flew seventy times and in 2000 only managed five flights. Far from having a failure rate of 1 in 10,000 it proved (unhappily) to have a failure rate closer to 1 in 25 (although by now it has become 1 in more than 100). Interestingly this is very close to the historic failure rate for solid rockets. The STS has taken off on schedule less than 50% of the time, and it costs $3 billion per year whether it flies or not (primarily for the standing army of support personnel). It was supposed to be frequent, cheap and manned. Instead, it is occasional, expensive and manned.

Flawed Space Policy

A fundamental difference with the Apollo experience is in the space policy which drove the Shuttle. Apollo had a clear simple goal, manon the moon within the decade. In contrast, the STS was all things to all people. It was initially conceived by NASA as the

A fundamental difference with the Apollo experience is in the space policy which drove the Shuttle. Apollo had a clear simple goal, man on the moon within the decade. In contrast, the STS was all things to all people. It was initially conceived by NASA as the "truck" which would carry humans and material to an Earth orbiting space station. It was also sold as the nation's primary launch system for all payloads, large and small. It was supposed to use the economics of reusability and be cheaper to fly than any existing or future expendable launch vehicle. It was to provide routine and frequent access to space. It was also to provide and carry orbiting lab facilities until a space station could be built. These were captured in the Reagan space policy of July 4, 1982 which defined the STS as the primary space launch system and said that it would be both fully operational and cost effective in providing routine access to space. The president also believed strongly in the commercialization of space, a policy that he tried with Landsat and foisted on the STS and NASA.

Since NASA wanted the STS to be primary US launch vehicle and wanted to justify the projected high flight rate it had to capture most of the launch market. Thus it got the Air Force to agree that all future military missiles would fly on the shuttle. The Air Force also agreed to refurbish the old Manned Orbiting Laboratory Space Launch Complex at Vandenberg AFB to have a site to launch into polar orbit from military missions. It of course required that all NASA payloads went on the shuttle. Thus the Hubble and Galileo were designed to go up on the Shuttle. It enticed the commercial customers in two ways. It offered very attractive prices for the first three years of Shuttle operations. Thus a PAM-D class satellite launch could be had for $15 million whereas to get the same launch in an Ariane was $30 million and $25 million on Delta. It also pulled it's payloads from Delta and Atlas. Since there were now being used less and less but they needed to sustain their infrastructure, their launch costs rose. Thus the Delta cost rose from $5 million a launch in 1970 to $26 million a launch by 1980. NASA also terminated the Delta and Atlas production lines in 1985. The Air Force did buy some Titan 34D's and contracted to buy only a few Titan 4's but did so over the objections of NASA and agreed to stop doing this. Thus NASA and the government moved to a one launcher policy driven by the desire for cost effectiveness. By January 1986, the STS had only flown twenty four times and had proven to be neither cheap nor reliable. However, so committed was NASA to the thesis that this was an operational vehicle that after only four test flights they had declared it an operational vehicle and on the 25th flight they were going to fly a teacher into space, an event to be watched by millions of schoolchildren. Instead of quick turnaround what they had found with this "operational" vehicle was that every one of the 17,000 tiles on it needed to be inspected after every flight and every SSME needed to be replaced every time. They had also noticed some worrisome erosion in the solid rocket joints where the segments were put together. Thus each Shuttle, instead of a turnaround of days, took months to prepare and required a large standing army of people to maintain it at human flight safety levels (0.99999). How could the 1977 estimates have been so wrong?

In retrospect, there were a number of factors. There was a deliberate NASA strategy of getting support for large programs with optimistic operational estimates and low cost estimates. This is the well known Camel's nose under the test strategy which basically relies on getting things going and building supporters who would sustain the program as the costs mounted. This strategy would be very clear on Station. In addition, the designers were overly optimistic about the technical progress of NASA. Perhaps they were still living in the glory days of Apollo. In any case they clearly underestimated the SSME difficulty. Still they seemed to have taken leave of common sense. The SSME is operated at 109% of total rated thrust. This is at the "red line". Any mechanic will tell you that an engine routinely operated at the "red line" will break down frequently. Truck engines (the model for the STS) work so reliably because they operate far from the maximum capabilities of the engine. The STS was certified as operational after only 4 flights with the really flight critical part, the ascent, being only 8 minutes each. Thus it was certified after 32 minutes of critical flight. In contrast the F-22 is required to be tested for a minimum of 183 hours of flight time before Congress authorizes buying the aircraft. Finally, the historical probability, based on many launches, of solid rocket failure has been 1 out of 25. How the NASA engineers managed to convince themselves that the catastrophic failure rate would be 1 in 10,000 when the STS had solid rockets on it, is hard to rationalize.

Primary Policy versus Secondary Policy

Another reason for the failure is in primary versus secondary policy. Primary policy breaks with past decisions and perspectives to meet the nation's top priorities. It has long term goals and has organized efforts to achieve them, so for Reagan primary policy was budget cuts, tax cuts and a huge defense buildup. For Bush primary policy was on the budget deficits. Primary policy is innovation. By contrast ancillary policy does not solve identified national problems. It has low grade status and receives limited attention and funding. Ancillary policy is the policy of continuation. By all these measures, in the 60's space policy was primary policy. It met the national angst after Sputnik and was bold and innovative. The Congress clearly bought in and money flowed freely (see Fig 8). There was broad public support and consensus on the goal, which was to show that the US could beat the Soviets. In contrast all the space policy behind the STS was secondary or ancillary policy. The interest in the space enterprise had declined in the public mind and there was no consensus between the White House and the Congress on where to go. There was no Vice President Johnson to build the consensus with the Congress. In primary policy the question is "What should we do?" In ancillary policy, the question becomes "What can we afford?" and "How can we sell it?" The STS and Space Station decision was marked by all of these large differences with the Apollo decision. The biggest and clearest way to see the difference between the two is to look at the difference in funding as a function of the Federal budget. This is a measure of the importance the administration and Congress really puts on something. In FY60, the NASA budget was 0.8% of the Federal budget. In FY66 it was 4.4% of the Federal budget, in FY80 it was back to 0.8% of the budget, in FY84 (Space Station) it was 0.8% of the budget and actually dropped the next year to 0.7% of the budget. In FY90 (Space Exploration Initiative) it was 0.99% of the budget and has since dropped significantly.

Effects of the Challenger disaster

The Challenger disaster struck the national psyche like Sputnik. It was made all the more visible by the fact that so many schoolchildren were watching because a teacher was on board. It plunged the space program and space policy into a huge crisis. Unhappily, there were several other launch failures that occurred at about the same time. These included in April 1986, a Titan 34D at Vandenberg and in May, a NASA Delta rocket that was launched into a thunderstorm. It seemed that NASA could do nothing right! The result was that all launch activity was grounded for several years while the technical issues were fixed & while the space policy was adjusted. The consequences of putting all the nation's eggs in only one major basket now meant that the US had no reliable means to get to space. The STS was grounded for 31 months and in that time space policy was transformed and the Air Force, commercial, international and NASA communities repositioned themselves. Since no launches were available on US rockets, many commercial satellite contractors turned to Arianespace. The US market shares of commercial launch plummeted and Ariane took significantly more than 50% of the free world market. In a sense, the seventies space policy of not allowing the French to use American rockets which pushed them to develop their own and putting all the US eggs in the Shuttle basket led directly to Ariane capturing most of the commercial market. Fortunately, many satellites had been designed to fly on the Shuttle and on the Ariane. After much debate in the space policy community, it was decided that the Shuttle would only be used for national security missions and for scientific missions where human presence was essential. All commercial communication satellites were pushed off the Shuttle and told to find other rides. This caused chaos in the commercial community and pushed them into the arms of Ariane. Of course this policy of using the Shuttle only when essential is a testament to the fact that it will never be an economic proposition.

The DoD decided that it wanted to move away from the Shuttle and return to a mixed fleet of expendable launch vehicles for assured access to space. Thus it cancelled the development of the Shuttle launch pad at Vandenberg AFB and restarted the Delta, Centaur and Titan lines. It agreed to buy 20 Deltas, 11 Centaurs and 24 Titans as a deliberate attempt by government policy to kick-start a dying industry. It also agreed to provide range support for all launches at the Cape Canaveral and Vandenberg AFB for only direct costs. Thus the DoD deliberately agreed to subsidize the commercial space industry, which was a form of space industrial policy.

The NASA scientific satellites were shelved to await the STS return to flights. Thus both Hubble and Galileo were put in storage to await later launch. In addition, the cryogenic Centaur upper stage for use from the Shuttle bay was cancelled. It was now seen as just too dangerous for a rare, high value asset like Shuttle. The direct consequence of this was that the Galileo mission when it flew would take two more years since there was now no upper stage to push it directly to Jupiter. In order to get there it would have to do a flyby past Venus and the Earth twice to get enough velocity. Since it is a radio isotope powered vehicle, this meant that 30 kg of plutonium came flying by the Earth twice to get to Jupiter. This has had the consequence of inflaming the anti-nuclear movement and probably set back substantially the use of nuclear power in space. The delays for Galileo and the Hubble turned out to have interesting consequences. For Hubble, it was fortunate since problems were discovered with the space telescope paint that would have been much harder to fix in orbit and may have limited its utility. For Galileo, it was bad. Galileo was shipped across the country three times (twice to the Cape and once back). This cross country trip and long storage led to the loss of lubricant in the high gain antenna which subsequently led to loss of that system on the way to Jupiter. Finally, NASA abandoned the policy of flying civilians (i.e. not regular astronauts) in the shuttle. In 1991, the President's advisory commission on space found the STS was still in the developmental phase after ten years of flights. So much for operational status!

Conclusion

The Challenger disaster illustrates the potential for coupled technology and policy decisions. It was a flawed technology choice driven by economic realities that coupled with a completely unrealistic set of policy statements set the stage for the disaster. It also showed the importance of clear communications and multiple ways to look at critical issues. Finally, it illustrated the nature of ethics in the face of bad decisions.

-

[1] For the concept of dynamic complexity and some characteristics of dynamically complex systems, we are indebted to John Sterman. For more information, see Business Dynamics: Systems Thinking and Modeling for a Complex World (Irwin/McGraw-Hill, 2000).