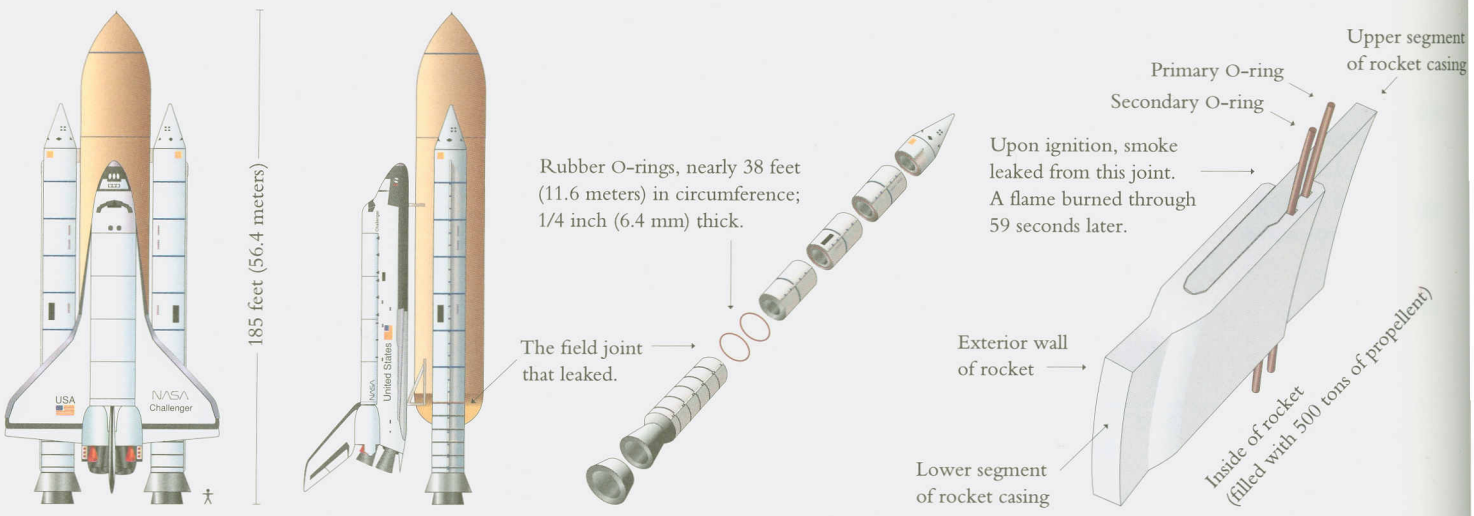
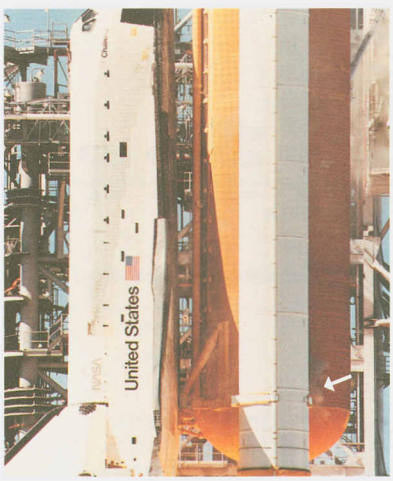
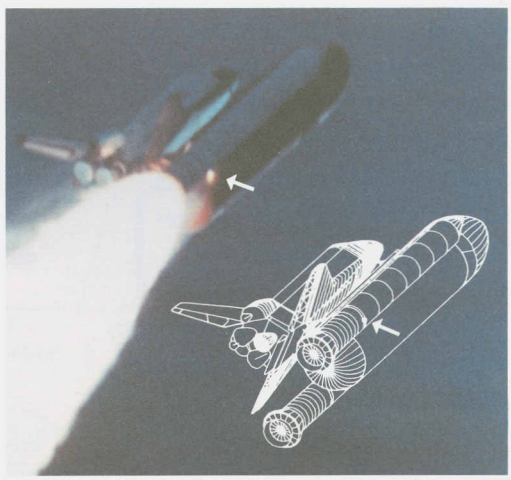


The shuttle consists of an *orbiter* (which carries the crew and has powerful engines in the back), a large liquid-fuel *tank* for the orbiter engines, and 2 solid-fuel *booster rockets* mounted on the sides of the central tank. Segments of the booster rockets are shipped to the launch site, where they are assembled to make the solid-fuel rockets. Where these segments mate, each joint is sealed by two rubber O-rings as shown above. In the case of the Challenger accident, one of these joints leaked, and a torch-like flame burned through the side of the booster rocket.



Less than 1 second after ignition, a puff of smoke appeared at the aft joint of the right booster, indicating that the O-rings burned through and failed to seal. At this point, all was lost.



On the launch pad, the leak lasted only about 2 seconds and then apparently was plugged by putty and insulation as the shuttle rose, flying through rather strong cross-winds. Then 58.788 seconds after ignition, when the Challenger was 6 miles up, a flicker of flame emerged from the leaky joint. Within seconds, the flame grew and engulfed the fuel tank (containing liquid hydrogen and liquid oxygen). That tank ruptured and exploded, destroying the shuttle.

As the shuttle exploded and broke up at approximately 73 seconds after launch, the two booster rockets crisscrossed and continued flying wildly. The right booster, identifiable by its failure plume, is now to the left of its non-defective counterpart.

The flight crew of Challenger 51-L. Front row, left to right: Michael J. Smith, pilot; Francis R. (Dick) Scobee, commander; Ronald E. McNair. Back row: Ellison S. Onizuka, S. Christa McAuliffe, Gregory B. Jarvis, Judith A. Resnik.

The Decision to Launch the Space Shuttle Challenger

On January 28, 1986, the space shuttle Challenger exploded and seven astronauts died because two rubber O-rings leaked.[22] These rings had lost their resiliency because the shuttle was launched on a very cold day. Ambient temperatures were in the low 30s and the O-rings themselves were much colder, less than 20°F.

One day before the flight, the predicted temperature for the launch was 26° to 29°. Concerned that the rings would not seal at such a cold temperature, the engineers who designed the rocket opposed launching Challenger the next day. Their misgivings derived from several sources: a history of O-ring damage during previous cool-weather launches of the shuttle, the physics of resiliency (which declines exponentially with cooling), and experimental data.[23] Presented in 13 charts, this evidence was faxed to NASA, the government agency responsible for the flight. A high-level NASA official responded that he was "appalled" by the recommendation not to launch and indicated that the rocket-maker, Morton Thiokol, should reconsider, even though this was Thiokol's only no-launch recommendation in 12 years.[24] Other NASA officials pointed out serious weaknesses in the charts. Reassessing the situation after these skeptical responses, the Thiokol managers changed their minds and decided that they now favored launching the next day. They said the evidence presented by the engineers was inconclusive, that cool temperatures were not linked to O-ring problems.[25]

Thus the *exact cause* of the accident was intensely debated during the evening before the launch. That is, for hours, the rocket engineers and managers considered the question: *Will the rubber O-rings fail catastrophically tomorrow because of the cold weather?* These discussions concluded at midnight with the decision to go ahead. That morning, the Challenger blew up 73 seconds after its rockets were ignited.

The immediate cause of the accident—an O-ring failure—was quickly obvious (see the photographs at left). But what are the general causes, the lessons of the accident? And what is the meaning of Challenger? Here we encounter diverse and divergent interpretations, as the facts of the accident are reworked into moral narratives.[26] These allegories regularly advance claims for the special relevance of a distinct analytic approach or school of thought: if only the engineers and managers had the skills of field X, the argument implies, this terrible thing would not have happened. Or, further, the insights of X identify the deep causes of the failure. Thus, in management schools, the accident serves as a case study for reflections about groupthink, technical decision-making in the face of political pressure, and bureaucratic failures to communicate. For the authors of engineering textbooks and for the physicist Richard Feynman, the Challenger accident simply confirmed what they already

[22] My sources are the five-volume *Report of the Presidential Commission on the Space Shuttle Challenger Accident* (Washington, DC, 1986) hereafter cited as *PCSSCA*; Committee on Science and Technology, House of Representatives, *Investigation of the Challenger Accident* (Washington, DC, 1986); Richard P. Feynman, "*What Do You Care What Other People Think?*" *Further Adventures of a Curious Character* (New York, 1988); Richard S. Lewis, *Challenger: The Final Voyage* (New York, 1988); Frederick Lighthall, "Launching the Space Shuttle Challenger: Disciplinary Deficiencies in the Analysis of Engineering Data," *IEEE Transactions on Engineering Management*, 38 (February 1991), pp. 63–74; and Diane Vaughan, *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA* (Chicago, 1996). The text accompanying the images at left is based on *PCSSCA*, volume I, pp. 6–9, 19–32, 52, 60. Illustrations of shuttle at upper left by Weilin Wu and Edward Tufte.

[23] *PCSSCA*, volume I, pp. 82–113.

[24] *PCSSCA*, volume I, p. 107.

[25] *PCSSCA*, volume I, p. 108.

[26] Various interpretations of the accident include *PCSSCA*, which argues several views; James L. Adams, *Flying Buttresses, Entropy, and O-Rings: The World of an Engineer* (Cambridge, Massachusetts, 1991); Michael McConnell, *Challenger: A Major Malfunction* (New York, 1987); Committee on Shuttle Criticality Review and Hazard Analysis Audit, *Post-Challenger Evaluation of Space Shuttle Risk Assessment and Management* (Washington, DC, 1988); Siddhartha R. Dalal, Edward B. Fowlkes, and Bruce Hoadley, "Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure," *Journal of the American Statistical Association*, 84 (December 1989), pp. 945–957; Claus Jensen, *No Downlink* (New York, 1996); and, cited above in note 22, the House Committee Report, the thorough account of Vaughan, Feynman's book, and Lighthall's insightful article.

knew: awful consequences result when heroic engineers are ignored by villainous administrators. In the field of statistics, the accident is evoked to demonstrate the importance of risk assessment, data graphs, fitting models to data, and requiring students of engineering to attend classes in statistics. For sociologists, the accident is a symptom of structural history, bureaucracy, and conformity to organizational norms. Taken in small doses, the assorted interpretations of the launch decision are plausible and rarely mutually exclusive. But when *all* these accounts are considered together, the accident appears thoroughly overdetermined. It is hard to reconcile the sense of inevitable disaster embodied in the cumulated literature of post-accident hindsight with the experiences of the first 24 shuttle launches, which were distinctly successful.

REGARDLESS of the indirect cultural causes of the accident, there was a clear proximate cause: an inability to assess the link between cool temperature and O-ring damage on earlier flights. Such a pre-launch analysis would have revealed that this flight was at considerable risk.[27]

On the day before the launch of Challenger, the rocket engineers and managers needed a quick, smart *analysis* of evidence about the threat of cold to the O-rings, as well as an effective *presentation* of evidence in order to convince NASA officials not to launch. Engineers at Thiokol prepared 13 charts to make the case that the Challenger should *not* be launched the next day, given the forecast of very chilly weather.[28] Drawn up in a few hours, the charts were faxed to NASA and discussed in two long telephone conferences between Thiokol and NASA on the night before the launch. The charts were unconvincing; the arguments against the launch failed; the Challenger blew up.

These charts have weaknesses. First, the title-chart (at right, where "SRM" means Solid Rocket Motor), like the other displays, does not provide the *names* of the people who prepared the material. All too often, such documentation is absent from corporate and government reports. Public, named authorship indicates responsibility, both to the immediate audience and for the long-term record. Readers can follow up and communicate with a named source. Readers can also recall what they know about the author's reputation and credibility. And so even a title-chart, if it lacks appropriate documentation, might well provoke some doubts about the evidence to come.

The second chart (top right) goes directly to the immediate threat to the shuttle by showing the history of eroded O-rings on launches prior to the Challenger. This varying damage, some serious but none catastrophic, was found by examining the O-rings from rocket casings retrieved for re-use. Describing the historical distribution of the *effect* endangering the Challenger, the chart does not provide data about the possible *cause*, temperature. Another impediment to understanding is that the same rocket has three different names: a NASA number (61A LH),

[27] The commission investigating the accident concluded: "A careful analysis of the flight history of O-ring performance would have revealed the correlation of O-ring damage and low temperature. Neither NASA nor Thiokol carried out such an analysis; consequently, they were unprepared to properly evaluate the risks of launching the 51-L [Challenger] mission in conditions more extreme than they had encountered before." *PCSSCA*, volume I, p. 148. Similarly, "the decision to launch STS 51-L was based on a faulty engineering analysis of the SRM field joint seal behavior," House Committee on Science and Technology, *Investigation of the Challenger Accident*, p. 10. Lighthall, "Launching the Space Shuttle," reaches a similar conclusion.

[28] The 13 charts appear in *PCSSCA*, volume IV, pp. 664–673; also in Vaughan, *Challenger Launch Decision*, pp. 293–299.

TEMPERATURE CONCERN ON
SRM JOINTS
27 JAN 1986

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View: Erosion Depth (in.)	Cross Sectional View: Perimeter Affected (deg)	Cross Sectional View: Nominal Dia. (in.)	Top View: Length Of Max Erosion (in.)	Top View: Total Heat Affected Length (in.)	Clocking Location (deg)
61A LH Center Field**	22A	None	None	0.280	None	None	36°--66°
61A LH CENTER FIELD**	22A	NONE	NONE	0.280	NONE	NONE	338°-18°
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
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.

Thiokol's number (SRM no. 22A), and launch date (handwritten in the margin above). For O-ring damage, six types of description (erosion, soot, depth, location, extent, view) break the evidence up into stupefying fragments. An overall index summarizing the damage is needed. This chart quietly begins to define the scope of the analysis: a handful of previous flights that experienced O-ring problems.[29]

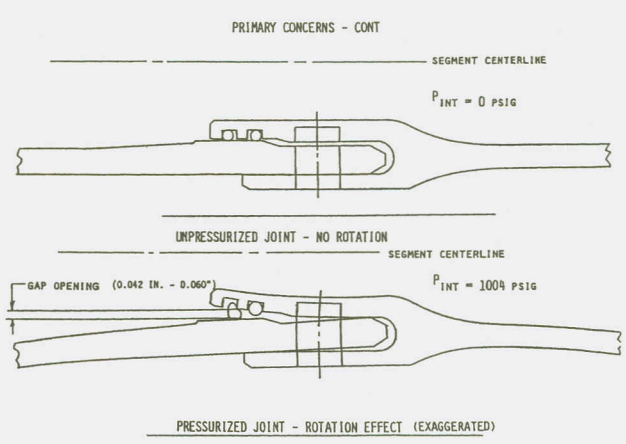
The next chart (below left) describes how erosion in the primary O-ring interacts with its back-up, the secondary O-ring. Then two drawings (below right) make an effective visual comparison to show how rotation of the field joint degrades the O-ring seal. This vital effect, however, is not linked to the potential cause; indeed, neither chart appraises the phenomena described in relation to temperature.

[29] This chart does not report an incident of field-joint erosion on STS 61-C, launched two weeks before the Challenger, data which appear to have been available prior to the Challenger prelaunch meeting (see *PCSSCA*, volume II, p. H-3). The damage chart is typewritten, indicating that it was prepared for an earlier presentation before being included in the final 13; handwritten charts were prepared the night before the Challenger was launched.

PRIMARY CONCERNS -

FIELD JOINT - HIGHEST CONCERN

- EROSION PENETRATION OF PRIMARY SEAL REQUIRES RELIABLE SECONDARY SEAL FOR PRESSURE INTEGRITY
  - IGNITION TRANSIENT - (0-600 MS)
    - (0-170 MS)HIGH PROBABILITY OF RELIABLE SECONDARY SEAL
    - (170-330 MS) REDUCED PROBABILITY OF RELIABLE SECONDARY SEAL
    - (330-600 MS) HIGH PROBABILITY OF NO SECONDARY SEAL CAPABILITY
- STEADY STATE - (600 MS - 2 MINUTES)
  - IF EROSION PENETRATES PRIMARY O-RING SEAL - HIGH PROBABILITY OF NO SECONDARY SEAL CAPABILITY
    - BENCH TESTING SHOWED O-RING NOT CAPABLE OF MAINTAINING CONTACT WITH METAL PARTS GAP OPENING RATE TO MEOP
    - BENCH TESTING SHOWED CAPABILITY TO MAINTAIN O-RING CONTACT DURING INITIAL PHASE (0-170 MS) OF TRANSIENT



BLOW BY HISTORY
SRM-15 WORST BLOW-BY
- 2 CASE JOINTS (80°), (110°) ARC
- MUCH WORSE VISUALLY THAN SRM-22

SRM 22 BLOW-BY
- 2 CASE JOINTS (30-40°)

SRM-13A, 15, 16A, 18, 23A 24A
- NOZZLE BLOW-BY

HISTORY OF O-RING TEMPERATURES (DEGREES-F)

MOTOR	MBT	AMB	O-RING	WIND
DM-4	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

Two charts further narrowed the evidence. Above left, "Blow-By History" mentions the two previous launches, SRM 15 and SRM 22, in which soot (blow-by) was detected in the field joints upon post-launch examination. This information, however, was already reported in the more detailed damage table that followed the title chart.[30] The bottom two lines refer to *nozzle* blow-by, an issue not relevant to launching the Challenger in cold weather.[31]

Although not shown in the blow-by chart, temperature is part of the analysis: SRM 15 had substantial O-ring damage and also was the coldest launch to date (at 53° on January 24, 1985, almost one year before the Challenger). This argument by analogy, made by those opposed to launching the Challenger the next morning, is reasonable, relevant, and weak. With only one case as evidence, it is usually quite difficult to make a credible statement about cause and effect.

If one case isn't enough, why not look at two? And so the parade of anecdotes continued. By linking the blow-by chart (above left) to the temperature chart (above right), those who favored launching the Challenger spotted a weakness in the argument. While it was true that the blow-by on SRM 15 was on a cool day, the blow-by on SRM 22 was on a *warm* day at a temperature of 75° (temperature chart, second column from the right). One engineer said, "We had blow-by on the hottest motor [rocket] and on the coldest motor."[32] The superlative "-est" is an extreme characterization of these thin data, since the total number of launches under consideration here is exactly *two*.

With its focus on blow-by rather than the more common erosion, the chart of blow-by history invited the rhetorically devastating—for those opposed to the launch—comparison of SRM 15 and SRM 22. In fact, as the blow-by chart suggests, the two flights profoundly differed: the 53° launch probably barely survived with significant *erosion* of the primary and secondary O-rings on both rockets as well as blow-by; whereas the 75° launch had no erosion and only blow-by.

[30] On the blow-by chart, the numbers 80°, 110°, 30°, and 40° refer to the *arc* covered by blow-by on the 360° of the field (called here the "case") joint.

[31] Following the blow-by chart were four displays, omitted here, that showed experimental and subscale test data on the O-rings. See *PCSSCA*, volume IV, pp. 664–673.

[32] Quoted in Vaughan, *Challenger Launch Decision*, pp. 296–297.

These charts *defined the database for the decision:* blow-by (not erosion) and temperature for two launches, SRM 15 and SRM 22. Limited measure of effect, wrong number of cases. Left out were the other 22 previous shuttle flights and their temperature variation and O-ring performance. A careful look at such evidence would have made the dangers of a cold launch clear. Displays of evidence implicitly but powerfully define the scope of the relevant, as presented data are selected from a larger pool of material. Like magicians, chartmakers reveal what they choose to reveal. That selection of data—whether partisan, hurried, haphazard, uninformed, thoughtful, wise—can make all the difference, determining the scope of the evidence and thereby setting the analytic agenda that leads to a particular decision.

For example, the temperature chart reports data for two developmental rocket motors (DM), two qualifying motors (QM), two actual launches with blow-by, and the Challenger (SRM 25) forecast.[33] These data are shown again at right. What a strange collation: the first 4 rockets were test motors that never left the ground. Missing are 92% of the temperature data, for 5 of the launches with erosion and 17 launches without erosion.

Depicting bits and pieces of data on blow-by and erosion, along with some peculiarly chosen temperatures, these charts set the stage for the unconvincing conclusions shown in two charts below. The major recommendation, "O-ring temp must be $\geq 53°$F at launch," which was rejected, rightly implies that the Challenger could not be safely launched the next morning at 29°. Drawing a line at 53°, however, is a crudely empirical result based on a sample of size one. That anecdote was certainly not an auspicious case, because the 53° launch itself had considerable erosion. As Richard Feynman later wrote, "The O-rings of the solid rocket boosters were not designed to erode. Erosion was a clue that something was wrong. Erosion was not something from which safety could be inferred."[34]

[33] The table of temperature data, shown in full at left, is described as a "History of O-ring Temperatures." It is a highly selective history, leaving out nearly all the actual flight experience of the shuttle:

MOTOR	O-RING	
DM-1	47	Test rockets ignited on fixed horizontal platforms in Utah.
DM-2	52	
QM-3	48	The only 2 shuttle launches (of 24) for which temperatures were shown in the 13 Challenger charts.
QM-4	51	
SRM-15	53	
SRM-22	75	
SRM-25	29 27	Forecasted O-ring temperatures for the Challenger.

[34] Richard P. Feynman, "*What Do You Care What Other People Think?*" *Further Adventures of a Curious Character* (New York, 1988), p. 224; also in Feynman, "Appendix F: Personal Observations on the Reliability of the Shuttle," *PCSSCA*, volume II, p. F2. On the many problems with the proposed 53° temperature line, see Vaughan, *Challenger Launch Decision*, pp. 309–310.

CONCLUSIONS:

- TEMPERATURE OF O-RING IS NOT ONLY PARAMETER CONTROLLING BLOW-BY

  SRM 15 WITH BLOW-BY HAD AN O-RING TEMP AT 53°F
  SRM 22 WITH BLOW-BY HAD AN O-RING TEMP AT 75°F
  FOUR DEVELOPMENT MOTORS WITH NO BLOW-BY WERE TESTED AT O-RING TEMP OF 47° TO 52°F

  DEVELOPMENT MOTORS HAD PUTTY PACKING WHICH RESULTED IN BETTER PERFORMANCE

- AT ABOUT 50°F BLOW-BY COULD BE EXPERIENCED IN CASE JOINTS

- TEMP FOR SRM 25 ON 1-28-86 LAUNCH WILL BE 29°F 9AM
  38°F 2PM

- HAVE NO DATA THAT WOULD INDICATE SRM 25 IS DIFFERENT THAN SRM 15 OTHER THAN TEMP

RECOMMENDATIONS:

- O-RING TEMP MUST BE ≥ 53°F AT LAUNCH

  DEVELOPMENT MOTORS AT 47° TO 52°F WITH PUTTY PACKING HAD NO BLOW-BY
  SRM 15 (THE BEST SIMULATION) WORKED AT 53°F

- PROJECT AMBIENT CONDITIONS (TEMP & WIND) TO DETERMINE LAUNCH TIME

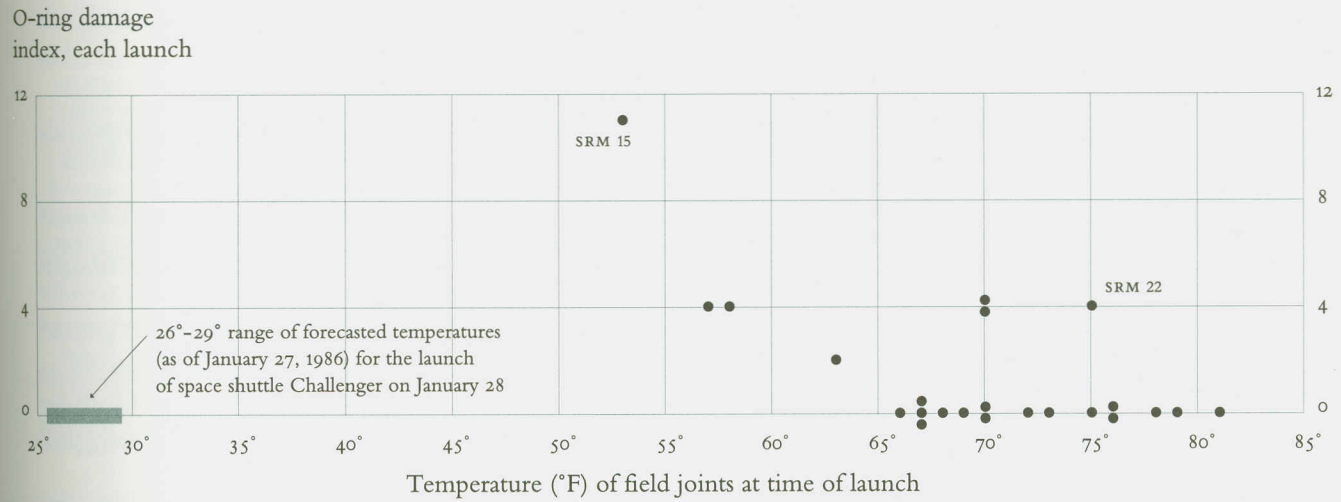
The 13 charts failed to stop the launch. Yet, as it turned out, the chartmakers had reached the right conclusion. They had the correct theory and they were thinking causally, but they were not *displaying* causally. Unable to get a correlation between O-ring distress and temperature, those involved in the debate concluded that they didn't have enough data to quantify the effect of the cold.[35] The displayed data were very thin; no wonder NASA officials were so skeptical about the no-launch argument advanced by the 13 charts. For it was as if John Snow had ignored some areas with cholera and *all* the cholera-free areas and their water pumps as well. The flights without damage provide the statistical leverage necessary to understand the effects of temperature. *Numbers become evidence by being in relation to.*

[35] *PCSSCA*, volume IV, pp. 290, 791.

This data matrix shows the complete history of temperature and O-ring condition for all previous launches. Entries are ordered by the possible cause, temperature, from coolest to warmest launch. Data in red were exhibited at some point in the 13 pre-launch charts; and the data shown in black were not included. I have calculated an overall O-ring damage score for each launch.[36] The table reveals the link between O-ring distress and cool weather, with a concentration of problems on cool days compared with warm days:

[36] For each launch, the score on the damage index is the severity-weighted total number of incidents of O-ring erosion, heating, and blow-by. Data sources for the entire table: *PCSSCA*, volume II, pp. H1-H3, and volume IV, p. 664; and *Post-Challenger Evaluation of Space Shuttle Risk Assessment and Management*, pp. 135–136.

Flight	Date	Temperature °F	Erosion incidents	Blow-by incidents	Damage index	Comments
51-C	01.24.85	53°	3	2	11	Most erosion any flight; blow-by; back-up rings heated.
41-B	02.03.84	57°	1		4	Deep, extensive erosion.
61-C	01.12.86	58°	1		4	O-ring erosion on launch two weeks before Challenger.
41-C	04.06.84	63°	1		2	O-rings showed signs of heating, but no damage.
1	04.12.81	66°			0	Coolest (66°) launch without O-ring problems.
6	04.04.83	67°			0	
51-A	11.08.84	67°			0	
51-D	04.12.85	67°			0	
5	11.11.82	68°			0	
3	03.22.82	69°			0	
2	11.12.81	70°	1		4	Extent of erosion not fully known.
9	11.28.83	70°			0	
41-D	08.30.84	70°	1		4	
51-G	06.17.85	70°			0	
7	06.18.83	72°			0	
8	08.30.83	73°			0	
51-B	04.29.85	75°			0	
61-A	10.30.85	75°		2	4	No erosion. Soot found behind two primary O-rings.
51-I	08.27.85	76°			0	
61-B	11.26.85	76°			0	
41-G	10.05.84	78°			0	
51-J	10.03.85	79°			0	
4	06.27.82	80°			?	O-ring condition unknown; rocket casing lost at sea.
51-F	07.29.85	81°			0	



When assessing evidence, it is helpful to see a full data matrix, all observations for all variables, those private numbers from which the public displays are constructed. No telling what will turn up.

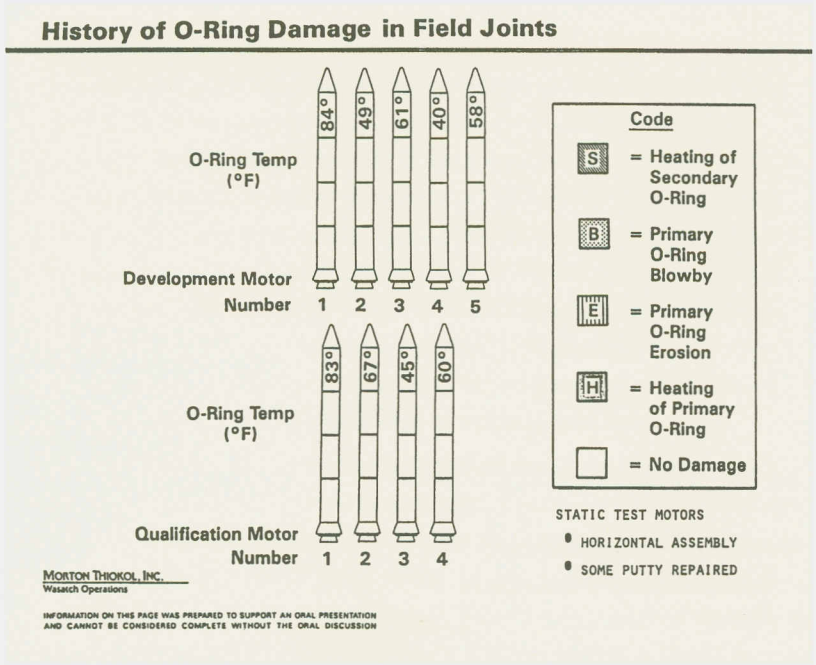
Above, a scatterplot shows the experience of all 24 launches prior to the Challenger. Like the table, the graph reveals the serious risks of a launch at 29°. Over the years, the O-rings had persistent problems at cooler temperatures: indeed, *every* launch below 66° resulted in damaged O-rings; on warmer days, only a few flights had erosion. In this graph, the temperature scale extends down to 29°, visually expressing the stupendous extrapolation beyond all previous experience that must be made in order to launch at 29°. The coolest flight without any O-ring damage was at 66°, some 37° warmer than predicted for the Challenger; the forecast of 29° is 5.7 standard deviations distant from the average temperature for previous launches. This launch was completely outside the engineering database accumulated in 24 previous flights.

In the 13 charts prepared for making the decision to launch, there is a scandalous discrepancy between the intellectual tasks at hand and the images created to serve those tasks. As analytical graphics, the displays failed to reveal a risk that was in fact present. As presentation graphics, the displays failed to persuade government officials that a cold-weather launch might be dangerous. In designing those displays, the chartmakers didn't quite know what they were doing, and they were doing a lot of it.[37] We can be thankful that most data graphics are *not* inherently misleading or uncommunicative or difficult to design correctly.

The graphics of the cholera epidemic and shuttle, and many other examples,[38] suggest this conclusion: *there are right ways and wrong ways to show data; there are displays that reveal the truth and displays that do not.* And, if the matter is an important one, then getting the displays of evidence right or wrong can possibly have momentous consequences.

[37] Lighthall concluded: "Of the 13 charts circulated by Thiokol managers and engineers to the scattered teleconferees, six contained no tabled data about either O-ring temperature, O-ring blow-by, or O-ring damage (these were primarily outlines of arguments being made by the Thiokol engineers). Of the seven remaining charts containing data either on launch temperatures or O-ring anomaly, *six of them included data on either launch temperatures or O-ring anomaly but not both in relation to each other.*" Lighthall, "Launching the Space Shuttle Challenger," p. 65. See also note 29 above for the conclusions of the shuttle commission and the House Committee on Science and Technology.

[38] Edward R. Tufte, *The Visual Display of Quantitative Information* (Cheshire, Connecticut, 1983), pp. 13–77.



SOON after the Challenger accident, a presidential commission began an investigation. In evidence presented to the commission, some more charts attempted to describe the history of O-ring damage in relation to temperature. Several of these displays still didn't get it right.[39]

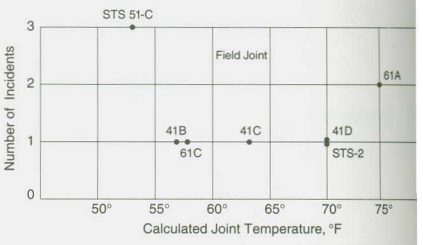
Prepared for testimony to the commission, the chart above shows nine little rockets annotated with temperature readings turned sideways. A legend shows a damage scale. Apparently measured in orderly steps, this scale starts with the most serious problem ("Heating of Secondary O-ring," which means a primary ring burned through and leaked) and then continues in several ordered steps to "No Damage." Regrettably, the scale's visual representation is disordered: the cross-hatching varies erratically from dark, to light, to medium dark, to darker, to lightest—a visual pattern unrelated to the substantive order of the measured scale. A letter-code accompanies the cross-hatching. Such codes can hinder visual understanding.

At any rate, these nine rockets suffered no damage, even at quite cool temperatures. But the graph is not on point, for it is based on test data from "Development and Qualification Motors"—all fixed rockets ignited on horizontal test stands at Thiokol, never undergoing the stress of a real flight. Thus this evidence, although perhaps better than nothing (that's all it is better than), is not directly relevant to evaluating the dangers of a cold-weather launch. Some of these same temperature numbers for test rockets are found in a pre-launch chart that we saw earlier.

Beneath the company logotype down in the lower left of this chart lurks a legalistic disclaimer (technically known as a CYA notice) that says

PCSSCA, volume V, p. 895.

[39] Most accounts of the Challenger reproduce a scatterplot that apparently demonstrates the analytical failure of the pre-launch debate. This graph depicts only launches with O-ring damage and their temperatures, omitting all damage-free launches (an absence of data points on the line of zero incidents of damage):



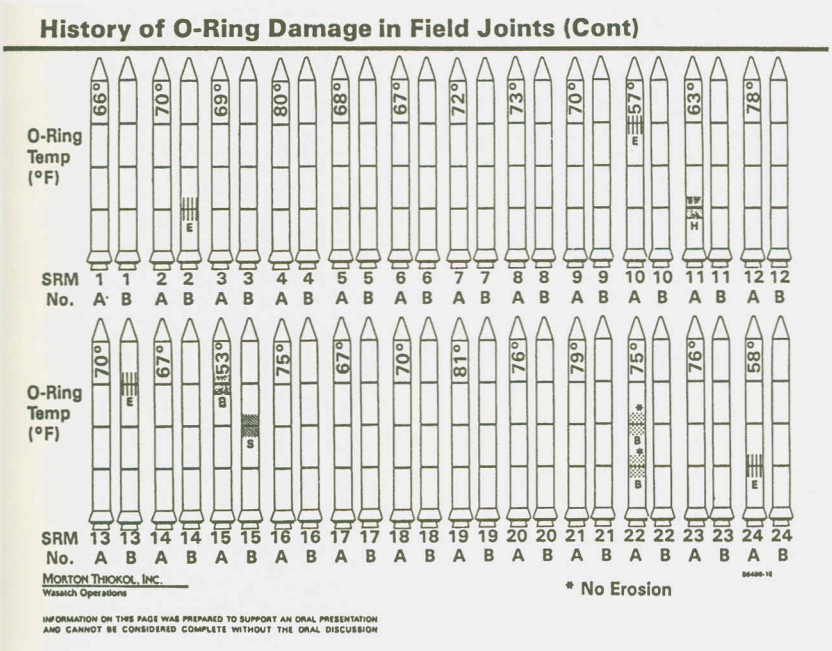
First published in the shuttle commission report (*PCSSCA*, volume I, p. 146), the chart is a favorite of statistics teachers. It appears in textbooks on engineering, graphics, and statistics—relying on Dalal, Fowlkes, Hoadley, "Risk Analysis of the Space Shuttle: Pre-*Challenger* Prediction of Failure," who describe the scatterplot as having a central role in the launch decision. (The commission report does not say when the plot was made.) The graph of the missing data-points is a vivid and poignant object lesson in how not to look at data when making an important decision. But it is too good to be true! First, the graph was *not* part of the pre-launch debate; it was *not* among the 13 charts used by Thiokol and NASA in deciding to launch. Rather, it was drawn *after* the accident by two staff members (the executive director and a lawyer) at the commission *as their simulation* of the poor reasoning in the pre-launch debate. Second, the graph implies that the pre-launch analysis examined 7 launches at 7 temperatures with 7 damage measurements. That is not true; only 2 cases of blow-by and 2 temperatures were linked up. The actual pre-launch analysis was much thinner than indicated by the commission scatterplot. Third, the damage scale is dequantified, only counting the number of incidents rather than measuring their severity. In short, whether for teaching statistics or for seeking to understand the practice of data graphics, why use an inaccurately simulated post-launch chart when we have the genuine 13 pre-launch decision charts right in hand? (On this scatterplot, see Lighthall, "Launching the Space Shuttle Challenger;" and Vaughan, *Challenger Launch Decision*, pp. 382–384.)

this particular display should not be taken quite at face value—you had to be there:

INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

Such defensive formalisms should provoke rambunctious skepticism: they suggest a corporate distrust both of the chartmaker and of any viewers of the chart.[40] In this case, the graph is documented in reports, hearing transcripts, and archives of the shuttle commission.

The second chart in the sequence is most significant. Shown below are the O-ring experiences of all 24 previous shuttle launches, with 48 little rockets representing the 24 flight-pairs:



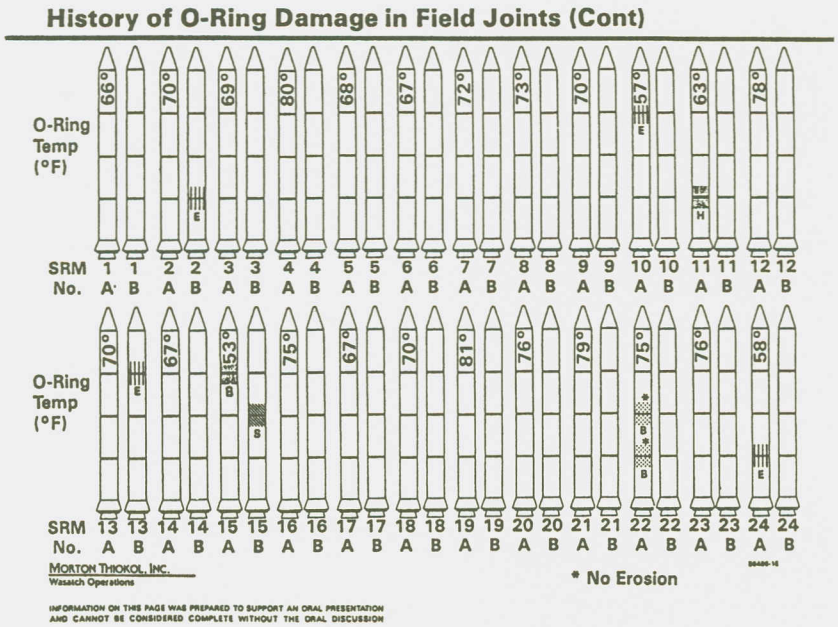
Rockets marked with the damage code show the seven flights with O-ring problems. Launch temperature is given for each pair of rockets. Like the data matrix we saw earlier, this display contains *all* the information necessary to diagnose the relationship between temperature and damage, if we could only see it.[41] The poor design makes it impossible to learn what was going on. In particular:

The Disappearing Legend At the hearings, these charts were presented by means of the dreaded overhead projector, which shows one image after another like a slide projector, making it difficult to compare and link images. When the first chart (the nine little rockets) goes away, the visual code calibrating O-ring damage also vanishes. Thus viewers need to memorize the code in order to assess the severity and type of damage sustained by each rocket in the 48-rocket chart.

[40] This caveat, which also appeared on Thiokol's final approval of the Challenger launch (reproduced here with the epigraphs on page 26), was discussed in hearings on Challenger by the House Committee on Science and Technology: "U. Edwin Garrison, President of the Aerospace Group at Thiokol, testified that the caveat at the bottom of the paper in no way 'insinuates . . . that the document doesn't mean what it says.'" *Investigation of the Challenger Accident*, pp. 228–229, note 80.

PCSSCA, volume v, p. 896.

[41] This chart shows the rocket pair SRM 4A, SRM 4B at 80°F, as having *undamaged* O-rings. In fact, those rocket casings were lost at sea and their O-ring history is unknown.



PCSSCA, volume v, p. 896. This image is repeated from our page 47.

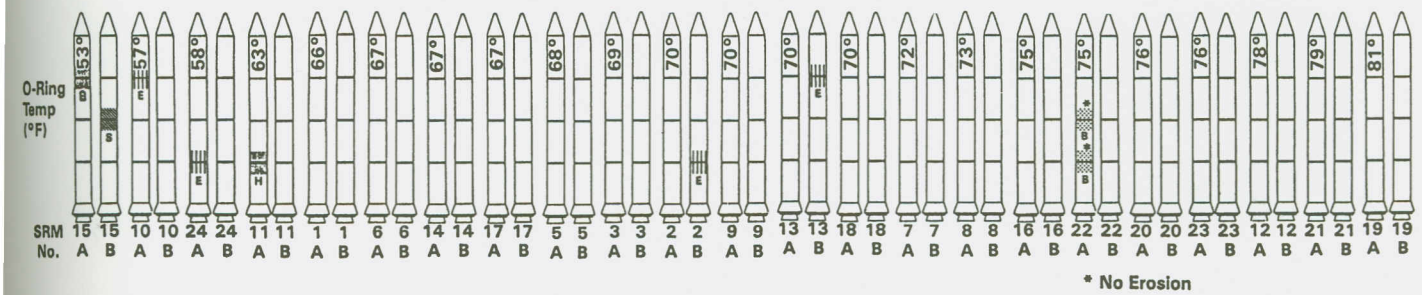
Chartjunk Good design brings *absolute attention* to data. Yet instead of focusing on a possible link between damage and temperature—the vital issue here—the strongest visual presence in this graph is the clutter generated by the outlines of the 48 little rockets. The visual elements bounce and glow, as heavy lines activate the white space, producing visual noise. Such misplaced priorities in the design of graphs and charts should make us suspicious about the competence and integrity of the analysis. Chartjunk indicates statistical stupidity, just as weak writing often reflects weak thought: "Neither can his mind be thought to be in tune, whose words do jarre," wrote Ben Jonson in the early 1600s, "nor his reason in frame, whose sentence is preposterous."[42]

Lack of Clarity in Depicting Cause and Effect Turning the temperature numbers sideways obscures the causal variable. Sloppy typography also impedes inspection of these data, as numbers brush up against line-art. Likewise garbled is the measure of effect: O-ring anomalies are depicted by little marks—scattered and opaquely encoded—rather than being totaled up into a summary score of damage for each flight. Once again Jonson's Principle: these problems are more than just poor design, for a lack of visual clarity in arranging evidence is a sign of a lack of intellectual clarity in reasoning about evidence.

Wrong Order The fatal flaw is the *ordering* of the data. Shown as a time-series, the rockets are sequenced by date of launching—from the first pair at upper left SRM No. 1A 1B to the last pair at lower right 24A 24B (the launch immediately prior to Challenger). The sequential order conceals the possible link between temperature and O-ring damage, thereby throwing statistical thinking into disarray. The time-series

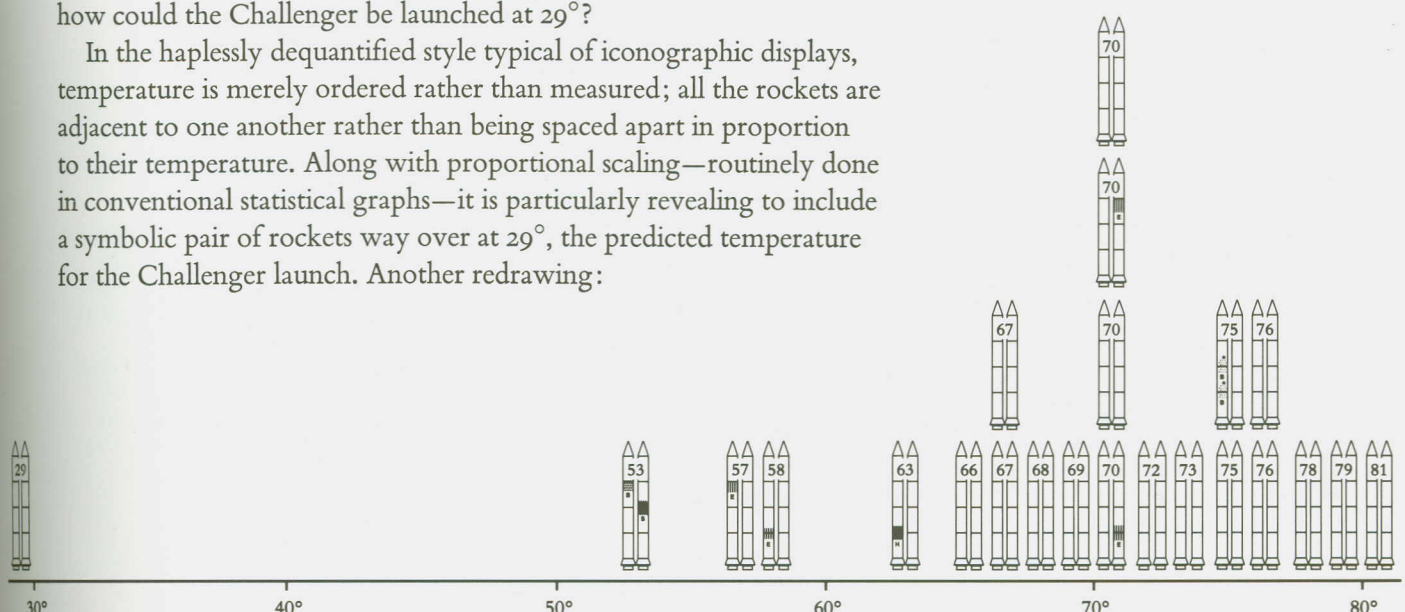
[42] Ben Jonson, *Timber: or, Discoveries* (London, 1641), first printed in the Folio of 1640, *The Workes* . . . , p. 122 of the section beginning with *Horace his Art of Poetry*. On chartjunk, see Edward R. Tufte, *The Visual Display of Quantitative Information* (Cheshire, Connecticut, 1983), pp. 106–121.

chart at left bears on the issue: Is there a time trend in O-ring damage? This is a perfectly reasonable question, but not the one on which the survival of Challenger depended. That issue was: Is there a temperature trend in O-ring damage?



Information displays should serve the analytic purpose at hand; if the substantive matter is a possible cause-effect relationship, then graphs should organize data so as to illuminate such a link. Not a complicated idea, but a profound one. Thus the little rockets must be *placed in order by temperature, the possible cause*. Above, the rockets are so ordered by temperature. This clearly shows the serious risks of a cold launch, for most O-ring damage occurs at cooler temperatures. Given this evidence, how could the Challenger be launched at 29°?

In the haplessly dequantified style typical of iconographic displays, temperature is merely ordered rather than measured; all the rockets are adjacent to one another rather than being spaced apart in proportion to their temperature. Along with proportional scaling—routinely done in conventional statistical graphs—it is particularly revealing to include a symbolic pair of rockets way over at 29°, the predicted temperature for the Challenger launch. Another redrawing:



Even after repairs, the pictorial approach with cute little rockets remains ludicrous and corrupt. The excessively original artwork just plays around with the information. It is best to forget about designs involving such icons and symbols—in this case and, for that matter, in nearly all other cases. These data require only a simple scatterplot or an ordered table to reveal the deadly relationship.

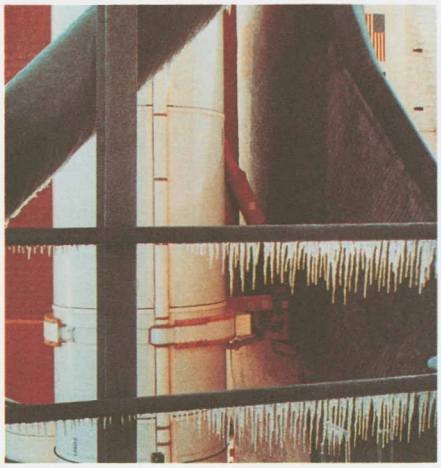
Photograph by Marilynn K. Yee, NYT Pictures, *The New York Times.*

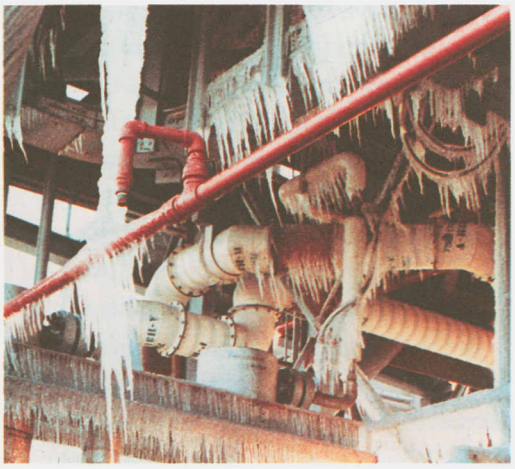
At a meeting of the commission investigating the shuttle accident, the physicist Richard Feynman conducted a celebrated demonstration that clarified the link between cold temperature and loss of resiliency in the rubber O-rings. Although this link was obvious for weeks to engineers and those investigating the accident, various officials had camouflaged the issue by testifying to the commission in an obscurantist language of evasive technical jargon.[43] Preparing for the moment during the public hearing when a piece of an O-ring (from a model of the field joint) would be passed around, Feynman had earlier that morning purchased a small clamp at a hardware store in Washington. A colorful theater of physics resulted. Feynman later described his famous experiment:

> The model comes around to General Kutyna, and then to me. The clamp and pliers come out of my pocket, I take the model apart, I've got the O-ring pieces in my hand, but I still haven't got any ice water! I turn around again and signal the guy I've been bothering about it, and he signals back, "Don't worry, you'll get it!". . . .
>
> So finally, when I get my ice water, I don't drink it! I squeeze the rubber in the C-clamp, and put them in the glass of ice water. . . .
>
> I press the button for my microphone, and I say, "I took this rubber from the model and put it in a clamp in ice water for a while."
>
> I take the clamp out, hold it in the air, and loosen it as I talk: "I discovered that when you undo the clamp, the rubber doesn't spring back. In other words, for more than a few seconds, there is no resilience in this particular material when it is at a temperature of 32 degrees. I believe that has some significance for our problem."[44]

[43] One official "gave a vivid flavor of the engineering jargon—the tang end up and the clevis end down, the grit blast, the splashdown loads and cavity collapse loads, the Randolph type two zinc chromate asbestos-filled putty laid up in strips—all forbidding to the listening reporters if not to the commissioners themselves." James Gleick, *Genius: The Life and Science of Richard Feynman* (New York, 1992), p. 422.

[44] Richard P. Feynman, "*What Do You Care What Other People Think?*" *Further Adventures of a Curious Character* (New York, 1988), pp. 151–153. Feynman's words were edited somewhat in this posthumously published book; for the actual hearings, see *PCSSCA*, volume IV, p. 679, transcript.





[45] Above, icicles hang from the service structure for the Challenger. At left, the photograph shows icicles near the solid-fuel booster rocket; for a sense of scale, note that the white booster rocket is 12 ft (3.7 m) in diameter. From *PCSSCA*, volume I, p. 113. One observer described the launch service tower as looking like ". . . something out of Dr. Zhivago. There's sheets of icicles hanging everywhere." House Committee on Science and Technology, *Investigation of the Challenger Accident*, p. 238. Illustration of O-ring experiment by Weilin Wu and Edward Tufte.

To create a more effective exhibit, the clamped O-ring might well have been placed in a transparent glass of ice water rather than in the opaque cup provided to Feynman. Such a display would then make a visual reference to the extraordinary pre-flight photographs of an ice-covered launch pad, thereby tightening up the link between the ice-water experiment and the Challenger.[45]

With a strong visual presence and understated conclusion ("I believe that has some significance for our problem"), this science experiment, improvised by a Nobel laureate, became a media sensation, appearing on many news broadcasts and on the front page of *The New York Times*. Alert to these possibilities, Feynman had intentionally provided a vivid "news hook" for an apparently inscrutable technical issue in rocket engineering:

> During the lunch break, reporters came up to me and asked questions like, "Were you talking about the O-ring or the putty?" and "Would you explain to us what an O-ring is, exactly?" So I was rather depressed that I wasn't able to make my point. But that night, all the news shows caught on to the significance of the experiment, and the next day, the newspaper articles explained everything perfectly.[46]

Never have so many viewed a single physics experiment. As Freeman Dyson rhapsodized: "The public saw with their own eyes how science is done, how a great scientist thinks with his hands, how nature gives a clear answer when a scientist asks her a clear question."[47]

[46] Feynman, "*What Do You Care What Other People Think?*", p. 153.

[47] Freeman Dyson, *From Eros to Gaia* (New York, 1992), p. 312.

AND yet the presentation is deeply flawed, committing the same type of error of omission that was made in the 13 pre-launch charts. Another anecdote, without variation in cause or effect, the ice-water experiment is *uncontrolled and dequantified*. It does not address the questions *Compared with what? At what rate?* Consequently the evidence of a one-glass exhibit is equivocal: Did the O-ring lose resilience because it was clamped hard, because it was cold, or because it was wet? A credible experimental

design requires at least two clamps, two pieces of O-ring, and two glasses of water (one cold, one not). The idea is that the two O-ring pieces are alike in all respects save their exposure to differing temperatures. Upon releasing the clamps from the O-rings, presumably only the cold ring will show reduced resiliency. In contrast, the one-glass method is not an experiment; it is merely an experience.

For a one-glass display, neither the cause (ice water in an opaque cup) nor the effect (the clamp's imprint on the O-ring) is explicitly shown. Neither variable is quantified. In fact, neither variable varies.

A controlled experiment would not merely evoke the well-known empirical connection between temperature and resiliency, but would also reveal the overriding *intellectual* failure of the pre-launch analysis of the evidence. That failure was a lack of control, a lack of comparison.[48] The 13 pre-launch charts, like the one-glass experiment, examine only a few instances of O-ring problems and not the causes of O-ring success. A sound demonstration would exemplify the idea that in reasoning about causality, *variations in the cause* must be explicitly and measurably linked to *variations in the effect*. These principles were violated in the 13 pre-launch charts as well as in the post-launch display that arranged the 48 little rockets in temporal rather than causal order. Few lessons about the use of evidence for making decisions are more important: story-telling, weak analogies, selective reporting, warped displays, and anecdotes are not enough.[49] Reliable knowledge grows from evidence that is collected, analyzed, and displayed with some good comparisons in view. And why should we fail to be rigorous about evidence and its presentation just because the evidence is a part of a public dialogue, or is meant for the news media, or is about an important problem, or is part of making a critical decision in a hurry and under pressure?

Failure to think clearly about the analysis and the presentation of evidence opens the door for all sorts of political and other mischief to operate in making decisions. For the Challenger, there were substantial pressures to get it off the ground as quickly as possible: an unrealistic and over-optimistic flight schedule based on the premise that launches were a matter of routine (this massive, complex, and costly vehicle was named the "shuttle," as if it made hourly flights from Boston to New York); the difficulty for the rocket-maker (Morton Thiokol) to deny the demands of its major client (NASA); and a preoccupation with public relations and media events (there was a possibility of a televised conversation between the orbiting astronaut-teacher Christa McAuliffe and President Reagan during his State of the Union address that night, 10 hours after the launch). But these pressures would not have prevailed over credible evidence against the launch, for many other flights had been delayed in the past for good reasons. Had the correct scatterplot or data table been constructed, no one would have dared to risk the Challenger in such cold weather.

[48] Feynman was aware of the problematic experimental design. During hearings in the afternoon following the ice-water demonstration, he began his questioning of NASA management with this comment: "We spoke this morning about the resiliency of the seal, and if the material weren't resilient, it wouldn't work in the appropriate mode, or it would be less satisfactory, in fact, it might not work well. I did a little experiment here, and *this is not the way to do such experiments*, indicating that the stuff looked as if it was less resilient at lower temperatures, in ice." (*PCSSCA*, volume IV, pp. 739–740, transcript, emphasis added.) Drawing of two-glass experiment by Weilin Wu and Edward Tufte.

[49] David C. Hoaglin, Richard J. Light, Bucknam McPeek, Frederick Mosteller, and Michael Stoto, *Data for Decisions: Information Strategies for Policymakers* (Cambridge, Massachusetts, 1982).

Conclusion: Thinking and Design

RICHARD Feynman concludes his report on the explosion of the space shuttle with this blunt assessment: "For a successful technology, reality must take precedence over public relations, for Nature cannot be fooled."[50] Feynman echoes the similarly forthright words of Galileo in 1615: "It is not within the power of practitioners of demonstrative sciences to change opinion at will, choosing now this and now that one; there is a great difference between giving orders to a mathematician or a philosopher and giving them to a merchant or a lawyer; and demonstrated conclusions about natural and celestial phenomena cannot be changed with the same ease as opinions about what is or is not legitimate in a contract, in a rental, or in commerce."[51]

In our cases here, the inferences made from the data faced exacting reality tests: the cholera epidemic ends or persists, the shuttle flies or fails. Those inferences and the resulting decisions and actions were based on various visual representations (maps, graphs, tables) of the evidence. The quality of these representations differed enormously, and in ways that governed the ultimate consequences.

For our case studies, and surely for the many other instances where evidence makes a difference, the conclusion is unmistakable: if displays of data are to be truthful and revealing, then the design logic of the display must reflect the intellectual logic of the analysis:

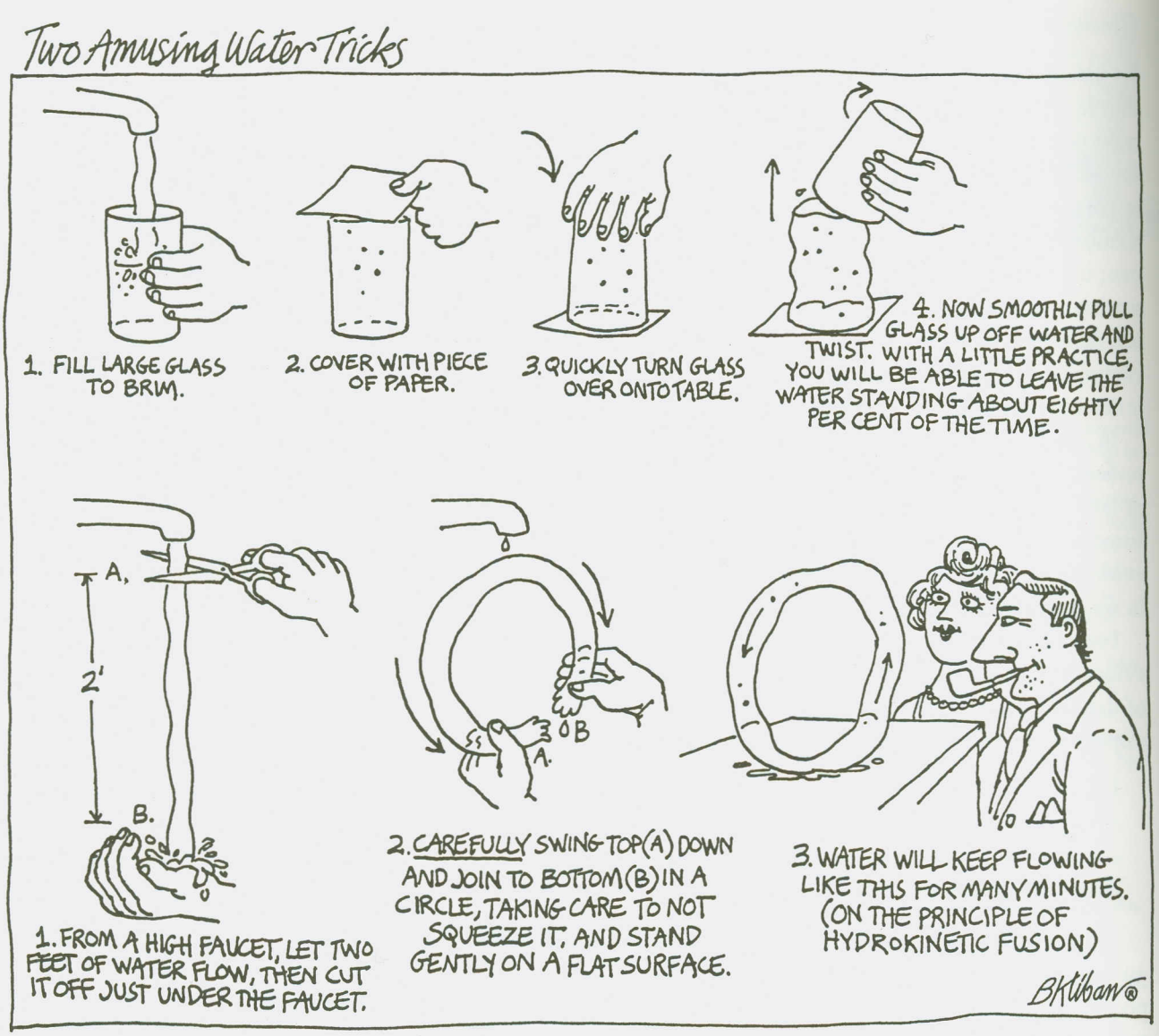
> *Visual representations of evidence should be governed by principles of reasoning about quantitative evidence. For information displays, design reasoning must correspond to scientific reasoning. Clear and precise seeing becomes as one with clear and precise thinking.*

For example, the scientific principle, *make controlled comparisons*, also guides the construction of data displays, prescribing that the ink or pixels of graphics should be arranged so as to depict comparisons and contexts. Display architecture recapitulates quantitative thinking; design quality grows from intellectual quality. Such dual principles—both for reasoning about statistical evidence *and* for the design of statistical graphics—include (1) *documenting* the sources and characteristics of the data, (2) insistently enforcing appropriate *comparisons*, (3) demonstrating mechanisms of *cause and effect*, (4) expressing those mechanisms *quantitatively*, (5) recognizing the inherently *multivariate* nature of analytic problems, and (6) inspecting and evaluating *alternative explanations*. When consistent with the substance and in harmony with the content, information displays should be documentary, comparative, causal and explanatory, quantified, multivariate, exploratory, skeptical.

And, as illustrated by the divergent graphical practices in our cases of the epidemic and the space shuttle, it also helps to have an endless commitment to finding, telling, and showing the truth.

[50] Richard P. Feynman, "Appendix F: Personal Observations on the Reliability of the Shuttle," *PCSSCA* volume II, p. F5; also, Feynman, "*What Do You Care What Other People Think?*" *Further Adventures of a Curious Character* (New York, 1988), p. 237.

[51] Galileo Galilei, letter to the Grand Duchess Christina of Tuscany, 1615, in *The Galileo Affair: A Documentary History*, edited and translated by Maurice A. Finocchiaro (Berkeley, 1989), p. 101.



B. Kliban, *Advanced Cartooning and Other Drawings* (New York, 1993), p. 25.