The Missing Term

The MISSING TERM

Correcting the Independence Assumption

Tal Ziskind

Rhombus Tech

Copyright © 2026 Rhombus Tech

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without the prior written permission of the publisher.

Published by Rhombus Tech

First edition, 2026

Contents

How to Read This Book 7

Prologue 10

I The Correction 12

1 The Error 13

2 The Crossover 21

II The Formulas 27

3 The Chain 28

4 The Update 36

5 The Grid 44

6 The Dose 51

7 The Building 58

8 The Redundancy 66

9 The Ensemble 73

Seven Formulas at a Glance **80**

III The Practice **82**

10 The Monitor **83**

11 The Trend **90**

12 The Barrier **95**

Epilogue **103**

Notes on Sources **106**

Key Formulas **109**

Glossary **110**

How to Read This Book

The idea in one paragraph. Almost every risk formula in use multiplies probabilities. If each of three backup systems has a 1% chance of failing, the chance that all three fail is $1\% \times 1\% \times 1\%$, one in a million. That arithmetic is right only if the systems fail for unrelated reasons. If they share a power supply, a software bug, a supplier, or a storm, the real chance is far higher: when the shared cause triples each system's chance of failing, all three failing together is twenty-seven times more likely than the formula says. This book takes seven formulas that institutions trust with lives and money, finds the multiplication sign in each, and puts the missing term back.

The Rogue Wave measured the error. It showed that any model treating coupled components as independent misprices compound events by a factor of d^n, where d is the measured dependence per component and n is the number of components that must fail together. It gave the forty-line test, ran it across twenty-nine datasets in twelve domains, and left the reader with a number: how wrong your model is.

This book corrects the formula.

Each domain in modern risk management has a governing equation—a formula that institutions trust, regulators enforce, and practitioners use daily. The procurement risk model counts suppliers. The grid reliability standard tests one failure at a time. The drug safety framework approves compounds individually. The building inspection schedule assumes structural elements degrade on their own. The flight safety case multiplies redundant-system reliabilities.

The ensemble forecast divides uncertainty by the number of models.

Every one of these formulas contains an independence assumption, and every one of them is wrong when the components are coupled: by a factor that grows as d^n when all the components must fail together, and by a smaller factor for the formulas that add risks or average estimates (Chapters 6 and 9). *The Rogue Wave* proved the first of these. This book writes the corrected version—formula by formula, domain by domain—and asks what would have been different had the correction been in place.

Three habits of Part II. Each of the seven domain chapters (Chapters 3 through 9) opens with the governing formula in its standard form, marks the independence assumption, and writes the corrected version. The correction is always in display math, boxed, so you can find it without rereading the chapter. Each closes with a counterfactual: given the historical failure that made the domain famous, what would the corrected formula have said beforehand? And each names the institutional barrier—the specific reason the correction has not been adopted—because a formula that nobody implements is not a fix. The opening scenes of Chapters 3, 4, and 6 are composites built to show the pattern; the events and figures cited in the text are documented in the Notes on Sources.

Where to start. Read Chapters 1 and 2 first; they carry the idea and the arithmetic. After that, the seven domain chapters of Part II stand on their own, so start with the one closest to your work and read the others in any order. If you would rather skip the equations, each domain chapter can be followed through its opening scene, *The Independence Assumption*, *The Counterfactual*, *The Barrier*, and the boxed summary at the end; the algebra in between supports those sections but is not needed to follow them. Sections that are mainly for specialists say so at their start. Part III is about using the correction day to day.

What you need. This book is self-contained but not self-starting.

Chapter 1 restates the d^n result in one chapter; a reader who has not read *The Rogue Wave* can follow the argument. A reader who has will recognize the result and can skim. The measurement code (the forty-line test) is not repeated here; it is in *The Rogue Wave*, Chapter 10, and available at the book's repository. Chapter 10 of this book gives the procedure in five steps, with a short sketch of the code.

Who this book is for. *The Rogue Wave* was written for anyone who uses a model. This book is written for anyone who writes one: the engineer drafting a reliability standard, the regulator setting a capital requirement, the procurement officer choosing a supplier portfolio, the pharmacist reviewing a medication list, the software architect drawing a dependency graph. It is an engineering companion to a scientific result.

What this book is not. It is not a survey of domains. *The Rogue Wave* already surveyed twelve. It is not a mathematical treatise; it uses the result and leaves the deeper theory aside. And it is not a software manual; the code is forty lines, published in *The Rogue Wave*, and not repeated here. This book is a repair manual. Seven formulas are broken. This book fixes them.

The sequence. This is the second book. *The Rogue Wave* presents the evidence and the law. *The Missing Term* corrects the formulas. A practitioner can read this book and go back to work.

Prologue

On the morning of March 23, 2021, the container ship *Ever Given* ran aground in the Suez Canal. Within days, at least 369 ships were queuing at both ends of the canal, and Lloyd's List estimated that about $400 million of goods were being delayed every hour—some $9.6 billion a day. The canal was blocked for six days.

One month earlier, Winter Storm Uri had shut down semiconductor fabrication plants in Austin, Texas—Samsung, NXP, Infineon—for weeks. At the same time, pandemic-driven demand for laptops, game consoles, and networking equipment had pushed global chip fabrication close to full capacity.

These three events—a grounding in Egypt, a freeze in Texas, and a virus from Wuhan—had no common physical cause. A cargo ship does not know about a polar vortex, and a virus does not read shipping schedules. The events were independent.

Their consequences were not.

All three hit the same supply chains, at the same time, through nodes that shared the same lean-inventory philosophy, the same geographic concentration, and the same just-in-time logistics that had been optimized for cost over decades. The automobile industry alone lost an estimated $210 billion in revenue in 2021 from semiconductor shortages. Global supply-chain disruptions persisted into 2023.

The procurement models that governed these supply chains counted suppliers. Six chip suppliers meant sixfold diversification. The models were not wrong about the number of suppliers. They

were wrong about what "diversification" means when every supplier draws from the same foundry, ships through the same strait, and runs on the same just-in-time buffer.

The error was not in the data. It was in the formula. The formula had a missing term.

This book is about that term—what it looks like in supply chains and in six other domains where the same term is missing from the same kind of formula. It is not a book about supply chains, or about software, or about power grids, or about drugs. It is a book about one correction, applied seven times. The correction is always the same: find the place in the formula where the model assumes independence, measure the coupling, and put the coupling back in.

In every domain, the formula works perfectly for one component. It fails exponentially for many coupled components. The missing term is the coupling.

Every formula in this book has the same shape: a probability raised to a power, or a sum of independent terms, or a product of uncorrelated errors. And every correction has the same shape: one measured number, inserted in one place, that changes the answer by a factor of two in the mildest cases and by orders of magnitude in the worst.

The number is d. The place is wherever the formula says "independent." The rest is arithmetic.

Part I

The Correction

Chapter 1

The Error

The question: How large is the pricing error when a model treats coupled components as independent, and why does it grow exponentially?

Every catastrophe in this book began with a multiplication sign.

A probability times a probability. A failure rate times a failure rate. A variance plus a variance. In each case, the arithmetic assumed that the terms were independent—that the thing on the left of the multiplication sign had nothing to do with the thing on the right. In each case, the terms were coupled, and the product was wrong.

This chapter restates a result from *The Rogue Wave* in one chapter, for the reader who begins here.

Every quantitative risk model rests on assumptions. The most common assumption, and the most consequential, is independence: the probability that component A fails and component B fails is the product of their individual failure probabilities.

$$P(A \text{ and } B) = P(A) \times P(B). \tag{1.1}$$

This is the multiplication rule for independent events. It is the first equation taught in every probability course, and it is embedded in every risk model that evaluates compound events—events in which multiple things must go wrong together.

When the components are actually independent, the formula is

exact. When they are coupled—when they share a common driver, a common supplier, a common environment, a common infrastructure—the formula understates the joint probability. *The Rogue Wave* showed that the understatement is not a small correction. It is exponential.

The Law

Define the *pricing ratio* R as the ratio of the real compound-event cost to the cost predicted by a model that treats the components as independent:

$$R = \frac{\text{real cost}}{\text{independent-model cost}}.$$

If $R = 1$, the model is correct. If $R = 100$, the model underpredicts by a factor of 100.

Across twenty-nine datasets spanning weather, hydrology, finance, commodities, housing, crypto, traffic, energy, epidemiology, seismology, agriculture, and chess, *The Rogue Wave* measured R and found that it follows a single law:

$$\boxed{R \approx d^n} \tag{1.2}$$

where $d > 1$ is the per-component dependence strength and n is the effective nonlinearity of the cost: the number of things that must go wrong together, counting both the components in distress and the periods they stay in distress. The exponent n is not a model parameter; it is the severity of the event you are pricing. Worse events have higher n. The error grows exponentially with severity. In the corrected formulas of Part II, the event is that all n components fail together, so n is the number of coupled components.

In plain terms, components are *coupled* when they share something that makes their failures arrive together: a supplier, a storm, a software bug, a power supply. The number d says how much that sharing multiplies each component's chance of failing in a compound

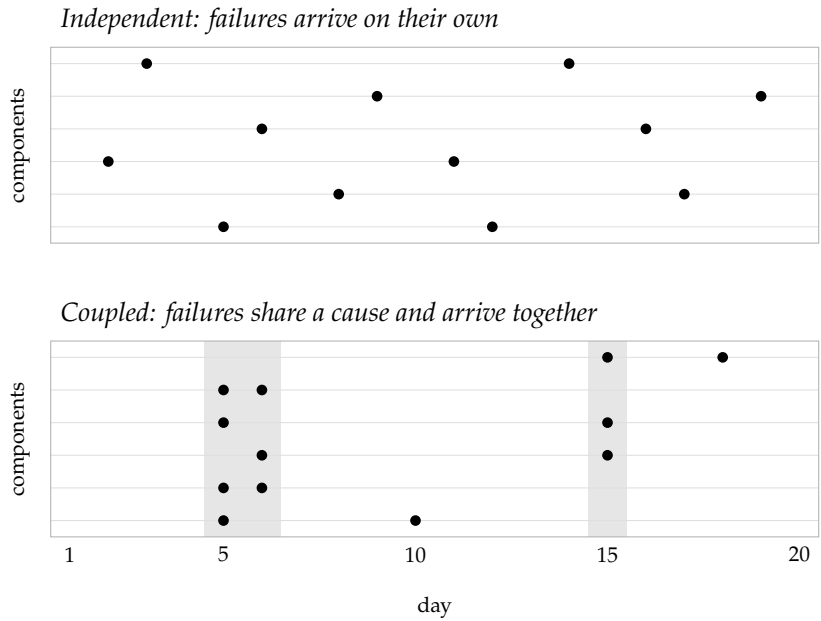

Figure 1.1: Twelve failures among six components over twenty days, drawn twice. Top: the failures arrive independently, and never more than one component is down on any day. Bottom: the same total number of failures, but the components share a cause, so failures arrive together; the shaded days have three or more components down at once. An independent model with the same overall failure rate (10%) gives both pictures the same forecast: about a third of one such day in twenty.

event. If $d = 1$, nothing is shared and the independent model is right. If $d = 2$, each component is twice as likely to fail as the model assumes. The event needs n components to fail together, so the doubling happens n times and the model is off by 2^n: eightfold for three components, 1,024-fold for ten. That is the whole law. Every chapter that follows is this idea applied to a different formula. Figure 1.1 shows what the sharing looks like in data.

The per-component dependence d varies by domain; *The Rogue Wave* puts it typically between 2 and about 15. Its pooled fit across

five domains gives each additional required day of a streak roughly a factor of 3 ($e^{1.15} \approx 3.2$ per day) and each additional required station about 1.3; the Texas running example alone runs near 11 per day. The $d \approx 3$ used in several worked examples in this book is the pooled per-day figure, taken as a round illustrative value. These values are estimated from the data by the permutation test, which shuffles the data to destroy one type of dependence at a time and measures how much of the real cost is lost.

The Permutation Test

The test requires three inputs: a matrix of observations (rows are time periods, columns are components), a cost function (how much damage when M or more components are simultaneously in distress), and a set of shuffles.

In plain terms, the test asks: if I scramble the data so that one kind of togetherness is destroyed, how much of the real-world damage disappears? The damage that disappears is what that kind of dependence was costing.

Three shuffles, each destroying one type of dependence (Figure 1.2):

1. **Column shuffle.** Independently permute each column. This preserves each component's own behavior (how often it is in distress, and how badly) but destroys everything else: the components can no longer move together, and no component's bad days can cluster in time. The ratio of real cost to column-shuffled cost is the *independence ratio*.

2. **Row shuffle.** Permute the rows (time periods) as whole units. This preserves which components move together on any given day but destroys temporal persistence: bad days can no longer follow bad days. The ratio is the *persistence ratio*.

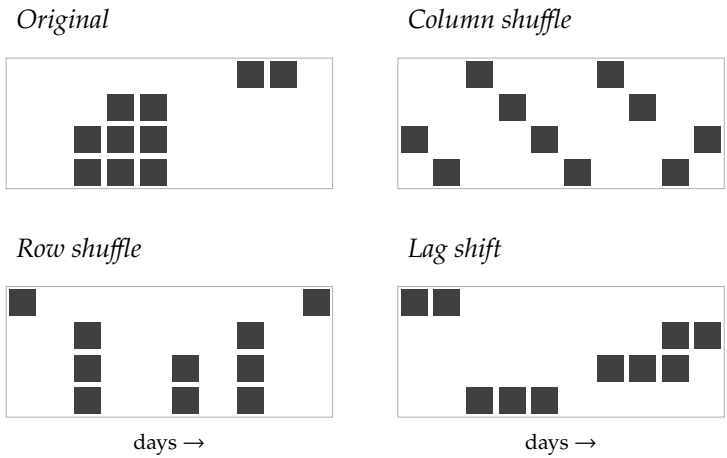

Figure 1.2: The three shuffles on a small matrix: four components (rows) over ten days (columns), with cold days filled in. *Original:* a cold spell strikes components 1 to 3 together on days 3 to 5. *Column shuffle:* each component's cold days are scattered at random, which destroys both the togetherness and the streaks. *Row shuffle:* whole days are reordered, so cold days still strike several components at once, but the spell is broken up in time. *Lag shift:* each component's series is slid by its own random amount, so its streaks survive but the components fall out of step.

3. **Lag shift.** Shift each column by a random lag (with wrap-around). This preserves each component's own streakiness (the tendency of a bad day to follow a bad day) but destroys synchronization: the components keep their individual rhythms but lose their alignment. The ratio is the *synchronization ratio*.

The total pricing ratio splits into a persistence part and a spatial part:

$$R_{\text{total}} \;=\; R_{\text{persistence}} \;\times\; R_{\text{spatial}}. \tag{1.3}$$

Here R_{spatial} is whatever the column shuffle destroys that the row shuffle does not: $R_{\text{spatial}} = R_{\text{total}}/R_{\text{persistence}}$. Because the second factor is defined as the remainder, the identity is exact, and on a log

scale the two parts simply add. (*The Rogue Wave* also reaches the same total along the other side of its decomposition diamond, as $R_{\text{sync}} \times R_{\text{own}}$.) The split tells you not just how wrong the model is, but *why*—how much of the error comes from events lasting too long (persistence) and how much from events hitting many components at once (spatial coupling).

The lag-shift ratio R_{sync} in the table below isolates timing alignment specifically. It is measured separately and is not the same quantity as R_{spatial}, so the last two columns of the table do not multiply to R.

A Pencil Example

The shuffles are easier to see on a matrix small enough to check by hand. Take two components and eight days. Each is in distress on three days, and they are the same three: days 3, 4, and 5. The event is that both are in distress on the same day, which happened on 3 of the 8 days.

Now run the column shuffle. Each component's three distress days land on any three of the eight days, independently. On any given day each component is in distress with probability $3/8$, so both are with probability $(3/8)^2 = 9/64$, and over eight days the expected count is $8 \times 9/64 = 9/8 = 1.125$ days. The real data has 3. The pricing ratio is $R = 3/1.125 = 8/3 \approx 2.7$. Two components must be in distress together, so $n = 2$ and $d = \sqrt{8/3} \approx 1.6$.

That is the entire procedure; real data has more rows and more columns. Shuffling the rows instead would change nothing here, because an event that lives inside a single day does not depend on the order of the days. Order matters only for streaks, which is why the persistence ratio grows with streak length.

The Scale of the Error

A few numbers from *The Rogue Wave* to fix the scale:

Domain	Streak	R	R_{pers}	R_{sync}
Texas cold (4 of 8 cities)	5 days	$\sim 10^5$	82×	31×
Equity drawdown (10 sectors)	5 days	~ 430	2.9×	113×
Potomac river flow	7 days	~ 890	388×	1.3×
Heavy rain (3 of 6 stations)	3 days	~ 590	3.9×	54×

Each row is a type of compound event. The first is five consecutive days on which at least four of eight Texas weather stations are at or below their own 10th-percentile temperature at once (their coldest tenth of days). R is how many times larger the real cost of that event is than the independent model predicts. R_{pers} and R_{sync} say how much of the gap comes from bad days following bad days and from components going bad in step.

The model is not wrong by 10% or even 100%. It is wrong by factors of hundreds to millions. These are not edge cases or adversarial constructions. They are the standard domains, the standard cost functions, and the standard data. The only nonstandard element is the test: nobody computed the ratio before, because nobody shuffled.

What This Means for a Formula

Every governing formula in this book—the procurement risk model, the grid reliability standard, the drug safety assessment, the structural inspection schedule, the flight safety case, the ensemble confidence interval—is a special case of the independence assumption. Each one multiplies, adds, or averages quantities that it treats as independent. Each one is wrong when the components are coupled, by d^n where all components must fail together.

The error is not in the model's parameters. It is in the model's structure. You cannot fix a d^n error by recalibrating. You fix it by

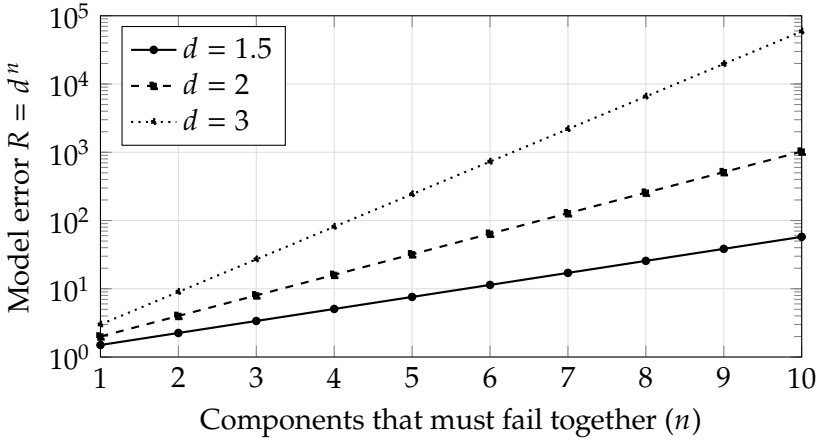

Figure 1.3: How far an independence-based model is off as more components must fail together, at three levels of coupling. The vertical axis is logarithmic: each gridline is ten times the one below it. At $d = 3$, ten coupled components make the model wrong by a factor of nearly sixty thousand.

adding the missing term.

The rest of this book adds it.

What to remember. Any model that treats coupled components as independent misprices compound events by approximately d^n. The error is structural, not parametric. You cannot calibrate it away. You must correct the formula.

Chapter 2

The Crossover

The question: When does adding components make the system less safe, not more?

Every risk manager knows the principle of diversification: spreading exposure across more components reduces risk. More suppliers mean less supply-chain risk. More generators mean more grid reliability. More drugs in a treatment protocol mean more therapeutic coverage. More models in an ensemble mean narrower uncertainty.

The principle is correct when the components are independent. When they are coupled, the principle has a boundary—and past it, diversification works in reverse.

In short: each added component helps less than the formula says, because the components share causes. Past a certain number, adding another component improves the formula's number by more than it improves the system.

The Arithmetic

Consider a system with n components, each with individual failure probability p. Under independence, the probability that all n fail simultaneously is:

$$P_{\text{IID}}(n) = p^n.$$

This drops exponentially with n. Adding components makes

simultaneous failure vanishingly unlikely. This is why we build redundant systems and diversified portfolios.

Under coupling with per-component dependence d, the corrected probability is:

$$P_{\text{coupled}}(n) \;=\; (d\,p)^n \;=\; d^n\,p^{\,n}.$$

This also drops with n—but only if $d \cdot p < 1$. And the *ratio* of real risk to modeled risk, which is d^n, rises exponentially. The model becomes exponentially more wrong as you add components, even if the absolute risk still falls.

But the absolute risk tells a subtler story. Consider not the probability of total failure but the expected number of simultaneous failures. Under independence, the expected number of components in distress on any given day is np. Under coupling, it rises to ndp. The risk per component is d times higher than the model predicts. Adding a component adds dp to the expected distress count, not p.

Now consider what diversification is supposed to buy. Under independence, each added component multiplies the chance of total failure by p, so the protection grows exponentially. Under coupling, each added component multiplies it by dp instead, while the model goes on assuming p. Every component you add therefore widens the gap between the protection the model claims and the protection you have by another factor of d. The next two sections turn this into two numbers: how many components it takes for the gap to pass a tolerance, and a measure of diversification that nets the coupling out.

The Crossover Point

For any coupling strength $d > 1$, there exists a number of components n^* beyond which the system's model error exceeds any desired threshold. The crossover satisfies:

$$\boxed{n^* \;=\; \frac{\ln T}{\ln d}} \tag{2.1}$$

where T is the maximum tolerable pricing error (e.g., $T = 10$ means you accept a tenfold underestimate). Beyond n^* components, the model is wrong by more than T. This is simply $d^n > T$ solved for n.

d	n^* at T=10	n^* at T=100	n^* at T=1000
1.2	12.6	25.3	37.9
1.5	5.7	11.4	17.0
2.0	3.3	6.6	10.0
3.0	2.1	4.2	6.3

Read the table row by row. At $d = 2$, the model is off by more than tenfold once the system has 3.3 components, by more than a hundredfold at 6.6, and by more than a thousandfold at 10.

At $d = 2$ (modest coupling), a system with more than 7 components has a pricing error exceeding 100×. At $d = 3$ (about the per-day factor *The Rogue Wave* measures for cold-weather streaks), a system with more than 4 components exceeds 100×.

This table is the reason the rest of this book exists. A procurement team with 6 suppliers and $d = 2$ is operating at $R = 64\times$. A grid with 5 coupled generators and $d = 3$ is at $R = 243\times$. A flight system with 3 coupled channels and $d = 10$ is at $R = 1000\times$. None of them know it, because the formula does not contain d.

The Diversification Trap

The crossover creates a trap. A risk manager who measures risk using the independent model sees risk falling as components are added. The model says: "You have 6 suppliers, your supply-chain risk is p^6, you are safe." The real risk is $d^6 p^6$, which is d^6 times higher. At $d = 1.8$, that is 34× higher. The risk manager has been rewarded for diversifying—promoted, given a larger budget, praised for best practice—while the real risk grew with every supplier added to a coupled system.

This is not a failure of risk management. It is a failure of the formula. The risk manager did exactly what the formula said to do. The formula was missing a term.

The same trap operates in every domain this book covers:

- **Supply chains.** Procurement teams add suppliers to reduce risk. If the suppliers are coupled (same geography, same shipping lane, same raw material), the pricing error grows with each addition.
- **Software.** Architects add microservices to isolate failures. If the services share infrastructure (same cloud region, same DNS, same certificate authority), the error grows with each service.
- **Power grids.** Planners add generators to increase reserve margin. If the generators share a fuel source or a weather exposure, the error grows with each generator.
- **Medicine.** Doctors add medications to address multiple conditions. If the drugs share a metabolic pathway, the interaction risk grows with each drug.
- **Aviation.** Engineers add redundant systems to reduce failure probability. If the systems share a failure mode (same software, same power bus, same environmental exposure), the error grows with each backup.
- **Climate science.** Modelers add ensemble members to narrow uncertainty. If the models share code heritage and calibration data, the effective sample size grows slower than n.
- **Construction.** Inspectors check more structural elements to increase confidence. If the elements share degradation drivers (same environment, same materials, same construction defects), the joint failure probability is higher than the product of individual probabilities.

In every case, the corrective is not to stop diversifying. It is to measure d and include it in the formula. Diversification works when d is low. It misleads when d is high and unaccounted for.

The Corrected Principle

The corrected diversification principle:

$$\boxed{\text{Effective diversification} \;=\; \frac{n}{d^{\,n-1}}} \tag{2.2}$$

This is a definition, not a derived law. It divides the nominal diversification, n, by the factor $d^{\,n-1}$: the model overstates its protection by $d^{\,n}$ for n components, of which the first component alone accounts for a factor of d. Read its peak as a rule of thumb for where adding components stops paying, not as a measured optimum.

When $d = 1$ (true independence), effective diversification equals n—the standard result. When $d > 1$, the denominator grows exponentially, and effective diversification peaks at a finite n^* and then declines. Past the peak, adding components makes the model worse, not better.

The peak occurs at $n_{\text{peak}} = 1/\ln d$. For $d = 1.5$: $n_{\text{peak}} \approx 2.5$. For $d = 2$: $n_{\text{peak}} \approx 1.4$. For $d = 3$: $n_{\text{peak}} \approx 0.9$—meaning the function is already declining at $n = 1$; any additional component makes the model worse. In words: the benefit of adding components rises at first and then falls, and the stronger the coupling, the sooner it falls.

At $d = 2$, you get maximum effective diversification from 1 to 2 components (the continuous peak is at $n \approx 1.4$; the integer values $n = 1$ and $n = 2$ tie). Every component past the second makes the model's error worse. This is not a recommendation to use only 2 suppliers. It is a statement that a model with 6 suppliers and no coupling term is more wrong than a model with 2 suppliers and no coupling term. The correct model has 6 suppliers *and* the coupling term.

A final note on limits. The corrected probability $(dp)^n$ is an

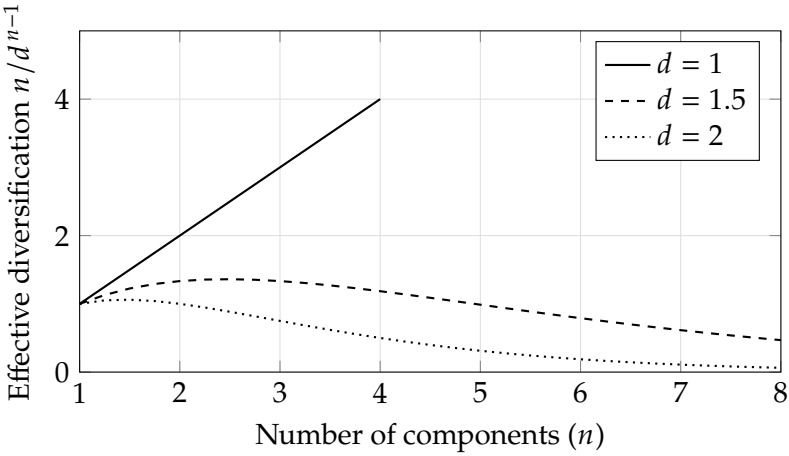

Figure 2.1: The benefit of adding components, net of coupling. With no coupling ($d = 1$) it equals the number of components (the line is cut off at $n = 4$). With coupling it rises, peaks, and falls, and the stronger the coupling, the lower and earlier the peak.

approximation; it is valid when $dp < 1$. When dp approaches or exceeds 1, the coupling is so strong that the shared driver dominates: all components fail together whenever the driver activates, regardless of n. In this regime—total coupling—the system's reliability is set by the shared driver's failure rate, and redundancy provides no benefit. Chapter 4 discusses this case in the context of shared software infrastructure.

What to remember. Diversification works when coupling is low. When coupling is high and unaccounted for, adding components makes the model exponentially more wrong. The crossover is at $n^* = \ln T / \ln d$ components. Past it, the formula needs the missing term.

Part II

The Formulas

Chapter 3

The Chain

The question: What is the corrected procurement risk formula, and what would it have said before the semiconductor shortage?

In January 2020, a procurement analyst at a major European automaker opened her quarterly risk dashboard and saw green across the board. Six semiconductor suppliers, spread across three countries, each rated low-risk by the vendor-scoring algorithm. The system counted suppliers the way a doctor counts vital signs: more is better, and six is safe.

By September 2021, all six were rationing supply to their customers—"on allocation," in the industry's phrase. Three had shut down entirely—one from a freeze in Texas, one from a COVID outbreak in Malaysia, one from a fire at a Renesas factory in Japan. The other three were running at capacity with 52-week lead times. The dashboard had turned red overnight, not because any single supplier had done anything unexpected, but because all of them had done the same thing at the same time.

The analyst's error was not in her data. It was in her formula.

A global supply chain is a network: raw materials flow through suppliers, fabricators, assemblers, and logistics providers to reach the final customer. At each tier, procurement teams choose suppliers. The standard risk assessment counts them.

The Governing Formula

The probability that a supply chain survives a disruption is modeled as one minus the probability that all suppliers fail:

$$P_{\text{survive}} \;=\; 1 - \prod_{i=1}^{n} q_i \tag{3.1}$$

where q_i is the probability that supplier i is disrupted and n is the number of suppliers. If all suppliers have the same disruption probability q, this simplifies to:

$$P_{\text{survive}} \;=\; 1 - q^n.$$

With $n = 6$ and $q = 0.05$ (a 5% annual disruption rate per supplier): $P_{\text{survive}} = 1 - 0.05^6 = 1 - 1.6 \times 10^{-8} \approx 1$. The model says the supply chain is essentially invulnerable.

The Independence Assumption

The product $\prod q_i$ is the independence assumption. It says: the event "supplier 1 is disrupted" and the event "supplier 2 is disrupted" are independent. The probability of both being disrupted is the product of their individual probabilities.

This is false whenever the suppliers share:

- **Geography.** A large share of the world's leading-edge semiconductor fabrication capacity is in Taiwan. A typhoon, an earthquake, or a geopolitical event affects all of them simultaneously. The suppliers are independent companies; their disruption probabilities are not independent variables.
- **Raw materials.** Roughly half of the world's semiconductor-grade neon gas—critical for chip lithography, the step that prints circuits onto silicon—came from two Ukrainian producers, in

Odesa and Mariupol, before the 2022 invasion. Every chip fabricator that depends on neon is coupled through the neon supply, regardless of where they are located.

- **Logistics.** The Strait of Malacca carries roughly a quarter of the world's seaborne trade. Suppliers on opposite sides of Asia share this chokepoint. A blockage or a security event couples them.

- **Inventory philosophy.** Just-in-time manufacturing minimizes inventory buffers. When a disruption arrives, every just-in-time node runs out simultaneously, because none of them has reserves. The shared absence of buffer stock is a coupling channel.

- **Demand.** In a demand surge (pandemic-driven electronics demand in 2020–2021), every supplier hits capacity simultaneously. The bottleneck is shared.

The Corrected Formula

The coupling between suppliers is captured by d—the per-component dependence that the permutation test estimates from supplier disruption data. The corrected formula replaces each supplier's disruption probability q with the coupled probability dq:

$$\boxed{P_{\text{survive}} \;=\; 1 - (d\,q)^n} \tag{3.2}$$

The same supply chain with $n = 6$, $q = 0.05$, and $d = 1.8$:

$$P_{\text{survive}} = 1 - (1.8 \times 0.05)^6 = 1 - (0.09)^6 = 1 - 5.3 \times 10^{-7}.$$

The corrected probability of total supply failure is 5.3×10^{-7}, compared to 1.6×10^{-8} under the independent model—a factor of $1.8^6 \approx 34$ higher. Still small in absolute terms.

But this is the probability of *all* six failing. The more relevant question is the probability of enough failing to disrupt production. If the supply chain needs at least 3 of 6 suppliers operational (a common threshold when each supplies a different component or geography), the probability that 4 or more are disrupted simultaneously is:

$$P(\geq 4 \text{ disrupted}) \;=\; \sum_{k=4}^{6} \binom{6}{k} (dq)^k (1 - dq)^{6-k}. \tag{3.3}$$

(The sum adds the chances that exactly 4, 5, or 6 of the suppliers are disrupted; $\binom{6}{k}$ counts the ways of choosing which k they are.)

Under independence ($d = 1$, $q = 0.05$): $P \approx 8.6 \times 10^{-5}$—negligible.

Under coupling ($d = 1.8$, $dq = 0.09$): $P \approx 8.5 \times 10^{-4}$—still small per year, but roughly 10× more likely.

Now consider a severe disruption—a pandemic, a war, a cascading logistics failure—where the base rate rises temporarily to $q = 0.3$:

Under independence: $P(\geq 4 \text{ disrupted}) \approx 0.070$.
Under coupling ($dq = 0.54$): $P(\geq 4 \text{ disrupted}) \approx 0.42$.

The model says a 7% chance. The corrected formula says 42%. During the 2021 semiconductor shortage, the observed outcome—sustained multi-node disruption across the global chip supply chain—was consistent with the corrected estimate.

Measuring d in Supply Chains

The permutation test requires a matrix. For supply chains:

- **Rows:** weeks or months.
- **Columns:** supply nodes (factories, ports, shipping routes, logistics hubs).
- **Entries:** disruption indicators—throughput as a fraction of capacity, delivery delays in days, or binary disruption flags.

Sources of data:

- **Shipping AIS data.** Automatic Identification System data tracks every commercial vessel in real time. Port congestion, route deviations, and sailing times are directly observable.
- **PMI subindices.** Purchasing Managers' Index surveys report supplier delivery times, order backlogs, and inventory levels monthly, by country.
- **Customs and trade flow databases.** UN Comtrade, national customs records, and industry-specific databases track volumes and delays.
- **Internal procurement data.** Every large buyer has delivery performance data for each supplier. This is the highest-resolution source and the least publicly available.

The test is the same forty lines from *The Rogue Wave*, applied to whichever matrix the organization can construct from its data. The column shuffle destroys the coupling between nodes; the ratio is d^n.

In the semiconductor supply chain, the coupling channels are strong. *The Rogue Wave* did not test supply chains empirically, and this book does not claim a measured d for chip suppliers. The worked examples in this chapter assume $d \approx 1.8$, an illustrative value for fabricators that share geography, neon supply, and shipping lanes; a firm with internal supplier data can compute d properly.

The Counterfactual

The automobile industry lost an estimated $210 billion in revenue in 2021 from the semiconductor shortage. What would the corrected formula have changed?

If procurement teams had computed d for their supplier portfolios and used the corrected formula, three things would have been different:

1. **Higher inventory buffers.** The corrected joint-disruption probability is 34× higher than the independent estimate. Cost-benefit analysis for safety stock—how much inventory to hold against disruption risk—would have justified larger buffers. Toyota's practice of holding semiconductor stockpiles since the 2011 earthquake is widely reported to have cushioned it in the early part of the 2021 shortage, though it too had to cut output sharply that September; as far as it went, it was empirically correct. The corrected formula would have made it analytically correct for every automaker.

2. **Diversification by coupling, not by count.** The corrected formula does not say "use fewer suppliers." It says "measure d between your suppliers and diversify to reduce d, not just to increase n." A procurement team with 6 suppliers in Taiwan has $n = 6$ but high d. Three suppliers on three continents has $n = 3$ but lower d. The corrected formula would favor the latter.

3. **Stress testing against coupled scenarios.** Standard supply-chain stress tests model the failure of one supplier at a time (the supply-chain analogue of the grid's N-1 criterion). The correction would require testing against d simultaneous failures: what happens when 2–3 suppliers are disrupted in the same quarter, which is d^n more likely than the independent model predicts?

Would $210 billion of losses have been prevented? No. The pandemic would still have caused a demand surge. The Suez Canal would still have been blocked. But the supply chains that faced these shocks would have had larger buffers, more geographic diversification, and a more accurate picture of their vulnerability. The industry's own post-mortem—which led to the CHIPS Act (2022), new fabrication capacity in the United States, Europe, and Japan, and industry-wide inventory increases—amounts to implementing

the correction after the fact. The corrected formula would have implemented it before.

The Barrier

The semiconductor shortage exposed the coupling. Why hasn't the correction been adopted?

- **Visibility.** Most companies cannot see past their tier-1 suppliers (the companies they buy from directly). To compute d, you need to know that your tier-1 suppliers share a tier-2 dependency—a company that they themselves buy from, such as the same foundry, the same chemical supplier, or the same shipping route. Sub-tier mapping is expensive and resisted by suppliers who consider their own supply chains proprietary.

- **Cost optimization.** The supply chain that maximizes coupling is also the supply chain that minimizes cost. Geographic concentration near ports, shared foundries that offer volume discounts, just-in-time inventory that frees working capital—every cost optimization increases d. The firm that corrects its formula must spend more on resilience. The firm that ignores the correction is cheaper, until the disruption arrives.

- **Competitive pressure.** In an industry where margins are thin and quarterly earnings drive equity prices, the firm that builds excess inventory, diversifies to more expensive suppliers, and stress-tests against coupled scenarios is at a competitive disadvantage every quarter that nothing goes wrong. The correction penalizes the early adopter.

- **Data standards.** No industry standard requires reporting of supplier coupling metrics. The $210 billion loss was attributed to "supply chain disruptions" in annual reports, not to "a 34× underestimate of joint failure probability due to unmeasured

coupling." Without standardized d reporting, the correction cannot be compared across firms or enforced by regulators.

The CHIPS Act addresses geographic concentration (one channel of coupling) but not the formula itself. Procurement risk models in 2026 still count suppliers.

What to remember. The procurement risk formula assumes independent supplier failures. The corrected formula replaces q with dq: the supply chain with 6 coupled suppliers at $d = 1.8$ is 34× more vulnerable than the model says. The correction calls for diversification by coupling, not by count.

Chapter 4

The Update

The question: What is the corrected service availability formula, and what would it have said before CrowdStrike?

The first call came in at 04:09 UTC on July 19, 2024, from an IT operations center in Sydney. Every Windows workstation in the building had blue-screened simultaneously. Rebooting produced another blue screen. By 04:30, the operations center had confirmed: it was not a local issue. By 05:00, they knew it was not an Australian issue. By 06:00, they knew it was not a regional issue.

CrowdStrike had pushed a channel file update to its Falcon endpoint security agent, security software installed on the computers of the organizations that use it. The update contained a logic error that crashed the Windows kernel—the core of the operating system, on which every other program depends. Within hours, 8.5 million Windows devices worldwide entered a boot loop—a blue screen of death on every restart.

Airlines cancelled more than 10,000 flights over the following days. Emergency services lost dispatch systems. Hospitals reverted to paper records. Banks could not process transactions. Supermarket point-of-sale systems went dark. Parametrix estimated direct losses of $5.4 billion for Fortune 500 companies alone.

Every affected organization had chosen its endpoint security vendor independently. They sat in different industries, different countries, different regulatory regimes. Their risk models—business con-

tinuity plans, disaster recovery assessments, vendor risk scorecards—treated each organization's IT infrastructure as independent.

They shared one kernel-level driver, pushed at the same time, to all of them.

The Governing Formula

Service-level agreements (SLAs) specify availability as a percentage of uptime:

$$A_{\text{system}} = \prod_{i=1}^{n} A_i \tag{4.1}$$

where A_i is the availability of component i and n is the number of components in series. If every component must work, the system is only as available as all of them together, so the availabilities multiply. "Five nines" ($A = 0.99999$) means at most 5.26 minutes of downtime per year; each additional "nine" cuts the downtime tenfold.

For components in parallel (redundant), the combined availability is:

$$A_{\text{redundant}} = 1 - \prod_{i=1}^{n}(1 - A_i) = 1 - f^n \tag{4.2}$$

where $f = 1 - A_i$ is the failure probability of each component. With two redundant components at $A = 0.999$: $A_{\text{redundant}} = 1 - 0.001^2 = 0.999999$—six nines from two three-nine components.

This is the standard calculation in every data-center design, every SLA negotiation, and every business-continuity plan. It assumes independent failures.

The Independence Assumption

The product $\prod(1 - A_i) = f^n$ says: the probability that component 1 fails *and* component 2 fails is the product of their individual failure probabilities. This requires that the components fail for independent

reasons.

In modern infrastructure, they do not. The coupling channels:

- **Shared cloud provider.** AWS hosts roughly 32% of cloud infrastructure. "Independent" services on the same cloud region share the same control plane (the management layer that starts, stops, and routes everything), the same network fabric, the same power grid, and the same software-defined networking. AWS us-east-1 outages in 2017, 2020, and 2021 each cascaded to hundreds of services that considered themselves independently hosted.

- **Shared software supply chain.** The CrowdStrike event demonstrated the extreme case: a single vendor's update, running at kernel level, on 8.5 million endpoints. Less extreme cases are ubiquitous: shared DNS providers (DNS is the internet's phone book, turning names into addresses; the 2016 Dyn attack took down Twitter, Reddit, Netflix, PayPal simultaneously), shared CDNs, the networks that deliver web content (the 2021 Fastly outage), shared certificate authorities, shared package registries (the 2016 left-pad incident).

- **Shared update timing.** Software updates are pushed to all customers simultaneously. A bug in version N hits every organization running version N at the same time. The update channel is a synchronization mechanism—it ensures that failures, when they occur, occur together.

- **Shared configuration.** Default configurations, best-practice templates, and compliance requirements create homogeneity. Organizations that independently follow the same best-practice guide end up with the same vulnerability surface.

The Corrected Formula

For redundant components with coupling d:

$$\boxed{A_{\text{redundant}} \;=\; 1 - (d \cdot f)^n} \tag{4.3}$$

Two "independent" backup systems with $f = 0.001$ and $d = 5$ (strong coupling through shared infrastructure):

$$A_{\text{redundant}} = 1 - (5 \times 0.001)^2 = 1 - 0.005^2 = 1 - 0.000025 = 0.999975.$$

The model promised six nines. The corrected formula gives four and a half. That is the difference between 32 seconds of downtime per year and 13 minutes—a factor of $d^n = 25$ in failure probability.

For the CrowdStrike event, the coupling was not $d = 5$. It was closer to total: every organization running the Falcon agent with auto-update received the faulty channel file. The effective d was bounded only by the vendor's market share within each sector. The "redundancy" of having separate IT teams, separate data centers, and separate business-continuity plans was irrelevant because the failure was injected below all of those layers, at the kernel.

Measuring d in Software Systems

The matrix:

- **Rows:** time intervals (minutes, hours, or days depending on resolution).
- **Columns:** services, instances, or organizations.
- **Entries:** availability indicators—response time, error rate, binary up/down status.

Data sources are abundant. Every modern software system generates telemetry: Prometheus metrics, Datadog dashboards, Cloud-

Watch logs, PagerDuty incident records. The permutation test can run on any of these.

Illustrative values for planning (assumptions, not published measurements):

- Services sharing a cloud region: $d \approx 2$ to 5 during regional incidents, $d \approx 1.1$ to 1.3 during normal operation.
- Services sharing a DNS provider: $d \approx 3$ to 8 during DNS incidents.
- Services sharing a software dependency: d can be effectively infinite during dependency failure (if the dependency is in the critical path, all services fail simultaneously—$d \cdot f = 1$).

The distinction between normal-operation d and incident d is important. Coupling is latent during normal operation and activates during stress. A system that measures d only during calm will underestimate its vulnerability. The correction must use the stress d, not the calm d.

This domain reveals something the other domains do not: the existence of *total coupling*, where $d \cdot f = 1$. When a shared dependency is in the critical path and it fails, every downstream service fails simultaneously, with probability 1. The coupling does not merely raise the probability; it saturates it. In this regime, redundancy is not weak—it is zero. Adding a second data center, a second network path, a second monitoring system provides exactly no protection against a failure in the shared layer. The corrected formula gives $(d \cdot f)^n = 1^n = 1$: certain failure regardless of n.

Total coupling is the extreme case, but partial-total coupling is common. Any infrastructure component shared by more than $1/f$ of your systems has $d \cdot f > 1$, meaning the corrected failure probability exceeds certainty for a single event—which in practice means the component *is* the system's reliability ceiling. The correction makes

this visible. The standard formula hides it behind a product of small numbers.

The Counterfactual

What would the corrected formula have changed before July 19, 2024?

The corrected SLA formula would have flagged any vendor with more than approximately $1/d$ market share in a critical category as a single point of failure, regardless of how many "independent" organizations used it. CrowdStrike's endpoint-security market share in the Fortune 500 was well above that threshold.

Three specific changes:

1. **Staged rollouts as a reliability requirement.** The correction shows that simultaneous deployment to all customers creates total coupling ($d \cdot f \approx 1$). A staged rollout—1% of endpoints first, then 10%, then 100% over days—reduces the effective d for the initial wave to near 1. The CrowdStrike update was pushed to 100% of endpoints in a single release. A staged-rollout requirement, derived from the corrected formula, would have limited the blast radius to the first cohort.

2. **Vendor diversification requirements.** Business-continuity standards (ISO 22301, SOC 2) assess vendor risk but do not measure coupling between an organization's vendor choices. The correction would require computing d across the vendor portfolio: if your endpoint security, your cloud provider, and your DNS all run on the same infrastructure stack, your "three independent layers" have $d \gg 1$ and your redundancy is illusory.

3. **Kernel-level access as a coupling multiplier.** Any software that runs at kernel level—endpoint security, drivers, hypervisors—has d equal to 1 for all applications on the same host: if the kernel

crashes, everything on that machine crashes. The correction makes this explicit: kernel-level software is not a component with an independent failure probability. It is a coupling channel with $d = 1/f$ (total coupling), and the system's reliability cannot exceed the kernel-level component's reliability, no matter how many application-level redundancies are in place.

The Barrier

- **The uptime illusion.** Cloud providers and SaaS vendors report uptime percentages that are computed per-service, not per-dependency-chain. AWS can report 99.99% uptime for S3 and 99.99% for EC2 and 99.99% for Route 53 while the actual availability of a service that depends on all three is lower—and the correlation between their outages is unaccounted for. Customers see three five-nine services and assume their combined reliability is fifteen nines.

- **Market concentration.** Three cloud providers (AWS, Azure, GCP) host the majority of internet infrastructure. Two CDNs (Cloudflare, Akamai) serve the majority of web traffic. A handful of endpoint-security vendors protect the majority of enterprise endpoints. The market structure creates coupling. The correction says the risk grows with market concentration, but antitrust law does not consider reliability coupling as a market-power concern.

- **Contractual structure.** SLAs are negotiated between a vendor and a customer. They do not cover the coupling between customers. A vendor's SLA with an airline does not account for the fact that many customers fail at the same moment, so that none of them can fall back on the others. The system failure was in the coupling, which no bilateral contract governs.

- **Speed over safety.** In cybersecurity, the advantage of rapid update deployment—patching a zero-day before it is exploited—conflicts with staged rollouts. The industry has optimized for speed because the threat model (adversaries moving fast) rewards it. The correction introduces a new cost: the probability of a correlated update failure grows with the speed and synchronization of deployment.

What to remember. The SLA availability formula assumes independent failures. The corrected formula replaces f with df: two "independent" backups sharing infrastructure with $d = 5$ give four and a half nines, not six. The correction calls for staged rollouts, vendor diversification by coupling, and treating kernel-level access as a reliability ceiling.

Chapter 5

The Grid

The question: What is the corrected grid reliability standard, and what would it have said before Uri?

At 1:25 a.m. on February 15, 2021, the frequency on the Texas grid was falling toward the point at which generating units trip offline. ERCOT later reported that the grid was four minutes and thirty-seven seconds from complete failure when load shedding began. A total collapse would have meant a black start—rebuilding the grid from nothing, plant by plant—which could take weeks.

The operator ordered rolling blackouts. They were not rolling. As natural gas pressure dropped—the electric compressor stations that pressurize the gas pipelines had themselves been shed—more generators tripped, which dropped more load, which shed more compressors. The feedback loop ran faster than the operators could react. Generation outages kept growing until they peaked at 52 gigawatts (GW), about the output of fifty large nuclear reactors and half the state's generating capacity. The blackouts lasted four days. In unheated homes across the state, pipes froze and burst. 246 people died.

The North American power grid is the largest machine ever built. Its reliability is governed by standards set by the North American Electric Reliability Corporation (NERC), and the most fundamental of those standards is the N-1 criterion: the system must be designed

to survive the loss of any single component—any one generator, any one transmission line, any one transformer—without shedding load.

The Governing Formula

The N-1 criterion is a deterministic standard, not a probabilistic one. It does not assign probabilities. It says: for every single component i, simulate its removal and verify that the remaining system can serve the load.

The probabilistic extension, used in resource-adequacy planning, computes the Loss of Load Expectation (LOLE):

$$\text{LOLE} \;=\; \sum_{k} P(k) \cdot \mathbb{1}[\text{capacity} - k < \text{load}] \tag{5.1}$$

where the sum is over all possible combinations of generator outages k, $P(k)$ is the probability of that combination, and $\mathbb{1}[\cdot]$ indicates whether the outage causes load shedding. In words: add up the chance of every combination of generator outages big enough to leave the grid short of power. The standard target is LOLE ≤ 0.1 days per year—one day of expected load shedding per ten years.

The probability $P(k)$ is computed from individual generator forced-outage rates, assumed independent:

$$P(\text{generators } i_1, \ldots, i_m \text{ all out}) \;=\; \prod_{j=1}^{m} q_{i_j} \tag{5.2}$$

where q_{i_j} is the forced-outage rate of generator i_j, the chance that it is unavailable because of an unplanned failure.

The Independence Assumption

The product $\prod q_{i_j}$ assumes that the outage of generator 1 is independent of the outage of generator 2. This is approximately true for

mechanical failures: a turbine blade fault in one generator does not cause a turbine blade fault in another.

It is approximately false for every other failure mode:

- **Weather.** Extreme cold reduces the output of gas plants (gas supply freezes, instrument lines freeze, cooling water freezes), wind turbines (blade icing, cut-out speeds), and solar panels (snow cover) simultaneously. Extreme heat increases demand (air conditioning) while reducing thermal plant efficiency and transmission capacity simultaneously. The weather is a common driver.

- **Fuel supply.** Gas-fired generators share gas pipelines. In Texas, 52% of generating capacity is gas-fired, and many gas plants share the same pipeline infrastructure. A pipeline failure or a supply curtailment affects all generators on that pipeline.

- **The gas–electric nexus.** Gas compressor stations that pressurize pipelines are powered by the electric grid. When the grid loses load and sheds customers, it may shed the compressor stations, which reduces gas supply, which causes more generators to trip, which sheds more load. The grid is coupled to its own fuel supply.

- **Market behavior.** In a wholesale electricity market, generators respond to the same price signals. When prices spike, every generator that can increase output does; when prices collapse, every generator that can shut down does. Market-driven correlations do not appear in engineering reliability models.

The Corrected Standard

Replace the N-1 criterion with N-d: the system must survive $\lceil d \rceil$ simultaneous component failures, where d is the measured coupling.

In probabilistic terms, the corrected LOLE replaces the independent outage probability with the coupled probability:

$$\boxed{P(\text{generators } i_1, \ldots, i_m \text{ all out}) \;=\; \prod_{j=1}^{m} (d \cdot q_{i_j})} \tag{5.3}$$

This raises the LOLE by a factor of d^m for any m-generator outage event. The N-1 criterion, which only considers $m = 1$, sees no change (the coupling factor for a single outage is just d, a rescaling of the outage rate). The error appears at $m \geq 2$: simultaneous outages of 2, 3, or more generators are d^m more likely than the independent model predicts.

For Texas during Winter Storm Uri:

The Rogue Wave measured about a factor of 3 per required day for cold streaks across its weather and river data, and about 11 per day for the Texas running example, but it did not measure generator outages. Generators share weather exposure *and* fuel-supply coupling, so a coupling of the same order is plausible. For illustration this chapter assumes $d_{\text{gen}} \approx 2.5$ for gas-dependent generators during extreme cold; it is an assumption, not a measurement.

With $d = 2.5$ and a typical forced-outage rate $q = 0.05$:

	Independent	Coupled ($d = 2.5$)
P(1 generator out)	0.050	0.125
P(2 generators out)	0.0025	0.016
P(3 generators out)	1.25×10^{-4}	1.95×10^{-3}
P(5 generators out)	3.1×10^{-7}	3.0×10^{-5}

The probability of 5 generators failing simultaneously is 97× higher under the coupled model. At 10 generators (the scale relevant to a grid crisis): $d^{10} = 2.5^{10} \approx 9{,}500\times$.

The Counterfactual

During Winter Storm Uri, ERCOT lost approximately 52 GW of generation capacity—roughly half the total installed capacity—in 48 hours. Hundreds of individual generating units tripped or derated. The standard LOLE model, which treats generator outages as independent, assigns an astronomically small probability to this event.

The corrected LOLE, with $d = 2.5$, shifts the probability of multi-generator outages by orders of magnitude. Specifically:

1. **The resource-adequacy assessment would have flagged winter risk.** A resource-adequacy assessment built on independent outage rates reports an adequate winter reserve margin. The corrected model, with weather-coupled outage rates, would have shown that the reserve margin during a cold snap could be negative: the expected coupled outage could exceed the margin.

2. **Winterization would have been cost-justified.** The cost-benefit analysis for winterization—insulating gas supply lines, installing cold-weather packages on turbines, weatherizing wind farms—depends on how likely the triggering event is. Suppose the independent model puts a statewide cold-weather generator failure at once in ten thousand years. A factor of about 100 (the five-generator ratio above) moves it to once in a hundred years. At that rate a Uri-scale loss of $195 billion costs about $2 billion a year in expectation, so winterization costing $5–15 billion once pays for itself within three to eight years.

3. **The gas–electric coupling would have been modeled.** The independent model treats gas supply as an external input. The corrected model recognizes that gas supply depends on the grid that it fuels. This circular dependency is a coupling channel with $d \gg 1$: when the grid fails, gas fails, which makes the grid

fail further. The corrected standard would require modeling this feedback loop explicitly.

There is a deeper lesson in the gas–electric nexus. In the other domains of this book, the correction is a static multiplier: measure d, multiply, done. In a feedback loop, d is not static. When the grid sheds load, it sheds compressor stations, which reduces gas pressure, which trips more generators, which sheds more load. Each cycle of the loop raises the effective d. The coupling is not a number; it is a process. A static correction ($p \to dp$) underestimates the risk of a feedback loop because d grows during the event. The corrected formula is still a better starting point than the uncorrected one—it gets the direction and the initial magnitude right—but for feedback systems, the full correction requires a dynamic model that updates d as the system degrades.

246 people died during Winter Storm Uri. The pricing error was in the formula.

The Barrier

- **The standard is slow to change.** NERC standards are developed through a consensus process involving utilities, regulators, system operators, and stakeholders. A change to the N-1 criterion—the most fundamental reliability standard in North America—requires years of technical review, comment periods, and regulatory approval. N-1 has been the foundation of North American planning standards since the aftermath of the 1965 Northeast blackout, which led to the creation of NERC.

- **Compliance infrastructure.** Every utility's planning tools, resource-adequacy software, and compliance documentation are built around independent outage rates. Switching to coupled rates requires new software, new data inputs (the coupling measurements), new calibration, and new validation. The

switching cost is borne by every utility; the benefit accrues to the system as a whole.

- **Cost allocation.** The corrected standard requires more reserve capacity, which costs money. In a deregulated market like Texas, the question of who pays for the additional reserves is unresolved. Generators profit from scarcity pricing during shortages; mandating additional reserves reduces scarcity, which reduces their revenue. The generators who would bear the cost have a financial interest in the independent model.

- **Weather coupling is getting worse.** Climate change is increasing the frequency of extreme weather events and expanding their geographic footprint. The coupling between generators—driven by weather—is rising. A corrected standard calibrated to historical d will be out of date within a decade. The standard must be dynamic, recomputed periodically from updated data, which is a further departure from the fixed N-1 criterion.

What to remember. The grid reliability standard assumes independent generator outages. The corrected standard replaces q with dq: during extreme weather with $d = 2.5$, five simultaneous outages are 97× more likely than the model says, and at ten generators, 9,500×. The correction requires modeling weather and fuel-supply coupling, not just mechanical failure.

Chapter 6

The Dose

The question: What is the corrected multi-drug safety assessment, and what would it have said about opioid–benzodiazepine co-prescription?

Mrs. Chen is 74 years old. She takes lisinopril for blood pressure, metformin for diabetes, atorvastatin for cholesterol, omeprazole for reflux, sertraline for depression, amlodipine for the blood pressure that lisinopril alone could not control, metoprolol because the amlodipine caused tachycardia, and now her rheumatologist wants to add meloxicam for knee pain.

Eight drugs. She sees four specialists, none of whom prescribed more than two. Each drug was approved in a trial where the typical participant was taking zero or one other medication. Each drug's safety profile—the list of side effects, the interaction warnings, the dosing guidelines—was established for the drug alone.

Nobody has established the safety profile of the combination.

A drug is approved by demonstrating that its benefits outweigh its risks in a controlled trial. The trial tests the drug against a placebo, in a population that is typically taking zero or one other medication. The drug's safety profile is established for the drug alone.

A patient takes the drug in combination with seven others.

The Governing Formula

The probability that a multi-drug regimen produces an adverse event is modeled—when it is modeled at all—as the union of individual drug risks, adjusted for known pairwise interactions:

$$P(\text{adverse event}) \;\approx\; 1 - \prod_{i=1}^{n}(1 - r_i) + \sum_{\text{known pairs}} \Delta_{ij} \tag{6.1}$$

where r_i is the adverse-event rate for drug i alone and Δ_{ij} is the known pairwise interaction between drugs i and j. In words: the chance that at least one drug causes trouble, treating each drug as acting alone, plus a patch for the pairs of drugs that are known to interact.

The product $\prod(1 - r_i)$ is the independence assumption: each drug's risk is independent of the others. The pairwise correction Δ_{ij} accounts for known two-drug interactions but not for three-drug, four-drug, or n-drug interactions.

The Independence Assumption

Drug interactions are not independent for the same reason that supply chains and power grids are not independent: the drugs share a common resource.

- **Metabolic pathways.** The cytochrome P450 (CYP) enzyme family, the liver's main system for breaking drugs down, metabolizes roughly 60 to 75% of marketed drugs. CYP3A4 alone handles about 30%. When two drugs compete for the same CYP enzyme, one inhibits the metabolism of the other, raising its blood concentration and its toxicity. When five drugs share CYP3A4, they form a five-way competition for the same enzyme. The interaction is not pairwise; it is combinatorial.

- **Organ systems.** Drugs that affect the same organ system—drugs that disturb the heart's rhythm (QT-prolonging), drugs that damage the kidney (nephrotoxic), drugs that damage the liver (hepatotoxic)—create additive or synergistic stress on that organ. The failure mode is shared: the organ, not the drug.

- **Physiological state.** An elderly patient with reduced kidney function, reduced liver mass, and reduced protein binding processes every drug differently from the healthy adults in the clinical trial. The physiological state couples the drugs by changing, all at once, how the body absorbs, processes, and clears each of them (their pharmacokinetics).

The Corrected Formula

For n drugs sharing a metabolic pathway, with per-drug adverse-event coupling d:

$$\boxed{P(\text{multi-drug adverse event}) \;=\; 1 - \prod_{i=1}^{n} (1 - d \cdot r_i)} \tag{6.2}$$

This replaces each drug's individual risk r_i with its coupled risk $d \cdot r_i$, reflecting that the presence of other drugs on the same pathway raises each drug's effective risk.

Example: a patient on 5 drugs sharing CYP3A4, each with $r_i = 0.02$ (a 2% adverse-event rate individually):

Under independence: $P = 1 - (1 - 0.02)^5 = 1 - 0.98^5 = 0.096 \approx 9.6\%$.

Under coupling with $d = 1.5$: $P = 1 - (1 - 0.03)^5 = 1 - 0.97^5 = 0.141 \approx 14.1\%$.

Under coupling with $d = 2.0$: $P = 1 - (1 - 0.04)^5 = 1 - 0.96^5 = 0.185 \approx 18.5\%$.

The absolute numbers are modest because r_i is small and n is

modest. But the relative increase is substantial: at $d = 2$, the multi-drug adverse-event rate is nearly double the independent estimate. For drugs with higher base rates (e.g., $r_i = 0.10$) or more drugs ($n = 8$), the gap widens further.

A necessary simplification. The flat multiplier d is a first-order approximation. Real pharmacokinetic coupling (how one drug changes another's level in the blood) is nonlinear: CYP inhibition is dose-dependent (a drug that is a mild inhibitor at 10 mg may be a strong inhibitor at 40 mg), saturable (the enzyme has a finite capacity, and competition becomes fiercer as more drugs load the pathway), and asymmetric (drug A may inhibit drug B's metabolism more than drug B inhibits drug A's). The pharmacodynamic coupling (how drugs add to one another's effects on the body) is worse: respiratory depression from opioids and benzodiazepines is synergistic, not additive—the combination is more dangerous than d times either drug alone.

The corrected formula captures the direction and the order of magnitude. It does not capture the pharmacokinetic detail. A patient-specific model would replace the flat d with a matrix of pairwise inhibition constants, derived from in-vitro enzyme kinetics. Such models exist in pharmacology (physiologically based pharmacokinetic models, or PBPK), but they are used in drug development, not in clinical practice. The flat d is the bridge: simple enough to implement in an electronic-health-record (EHR) alert, accurate enough to flag the patients who need the detailed model.

Measuring d in Pharmacology

The matrix:

- **Rows:** patients (or patient-time intervals).
- **Columns:** drugs.

- **Entries:** adverse-event indicators (binary: event/no-event during the observation period).

The data source is the FDA Adverse Event Reporting System (FAERS), which contains millions of reports of suspected drug adverse events, each listing the drugs the patient was taking. The permutation test shuffles the drug columns independently across patients, destroying the coupling between drugs while preserving each drug's marginal adverse-event rate.

No published estimate of d for drugs sharing a CYP enzyme is established here; the worked examples above assume d between 1.5 and 2, an illustrative range, with higher values plausible for strong inhibitors or inducers of the shared enzyme.

The Counterfactual

The opioid–benzodiazepine co-prescription crisis:

Between 2001 and 2013, the share of opioid users in the United States who also used a benzodiazepine rose from 9% to 17%, and concurrent users had 2.14 times the odds of an emergency visit or admission for opioid overdose (Sun et al.). Both drug classes cause respiratory depression. Together, they are synergistic: the respiratory depression from the combination exceeds the sum of the individual effects. The coupling channel is pharmacodynamic (both suppress the brainstem respiratory center) and pharmacokinetic (both compete for CYP3A4 metabolism).

In 2015, a case-cohort study of veterans receiving opioid analgesics found that current benzodiazepine prescription carried an adjusted hazard ratio of 3.86 for death from overdose (Park et al.): among veterans on opioids, those with a current benzodiazepine prescription died of overdose at 3.86 times the rate of those without one, after statistical adjustment for other differences. In 2016, the CDC issued a guideline recommending against concurrent prescribing. A near-

fourfold increase in risk from adding one coupled drug ($d \cdot r$ instead of r, with $d \approx 3.9$ for this specific combination at the overdose-death endpoint) is consistent with the framework: the interaction risk was not pairwise—it was driven by the shared physiological mechanism. The value $d \approx 3.9$ is read off the outcome, so this illustrates the framework rather than testing it.

The corrected formula, applied to the 2010 prescribing data, would have flagged opioid–benzodiazepine co-prescription as a high-coupling combination years before the epidemiological evidence accumulated. The measurement requires only FAERS data, not a prospective trial.

More broadly: the FDA's current framework reviews drug interactions one pair at a time. A drug label lists known interactions with specific other drugs. There is no systematic assessment of multi-drug coupling—the risk that arises when a patient takes 5, 8, or 12 drugs that share metabolic pathways. The corrected formula is a starting point for that assessment.

The Barrier

- **Regulatory structure.** The FDA approves drugs individually. Post-market surveillance monitors adverse events but does not systematically compute multi-drug coupling. Requiring a coupling assessment for polypharmacy would change the regulatory framework.

- **Trial design.** Testing all n-way drug combinations is combinatorially infeasible. A patient on 8 drugs has $\binom{8}{2} = 28$ pairs, $\binom{8}{3} = 56$ triples, and 163 combinations of four or more drugs. The corrected formula does not require testing each combination; it requires measuring d from observational data (FAERS). But observational data is considered lower-quality evidence than randomized trials.

- **Prescribing culture.** In clinical practice, each specialist prescribes within their domain. The cardiologist adds a statin, the psychiatrist adds an SSRI, the rheumatologist adds a biologic. No single physician sees the complete medication list in the context of metabolic coupling. Electronic health records list the drugs but do not compute d.

- **Polypharmacy is growing.** The fraction of US adults taking 5 or more prescription medications rose from 8.2% in 1999–2000 to 15% in 2011–2012 (Kantor et al., JAMA 2015), and it is higher among older adults. The population exposed to multi-drug coupling is growing, and the corrected formula says the risk grows faster than the number of drugs.

What to remember. Drug safety is assessed per drug, with pairwise interaction checks. The corrected formula replaces r_i with $d \cdot r_i$ for drugs sharing a metabolic pathway: at $d = 2$, a 5-drug regimen's adverse-event rate nearly doubles. The correction calls for multi-drug coupling assessment from observational data, not just pairwise interaction screening.

Chapter 7

The Building

The question: What is the corrected structural inspection model, and what would it have said about Champlain Towers South?

At 1:22 a.m. on June 24, 2021, a security camera in the parking garage of Champlain Towers South recorded the last seconds. A concrete slab sagged. Dust fell. Then the column beneath it buckled, and the slab above dropped onto the next column, which buckled, and the cascade ran through the east wing of the 12-story building in approximately 12 seconds. The camera feed ended. 98 people died.

The building had been inspected in 2018. It had a 40-year recertification due in 2021. The inspection process, mandated by Miami-Dade County, examines structural elements—columns, beams, slabs, connections—and assesses each element's condition.

This chapter is different from the others in this book. In the other domains, the coupling makes the failure more likely than the model predicts. In a building, the coupling changes the *mode* of failure: from a local crack (survivable, repairable) to a progressive collapse (catastrophic, irreversible). The independence assumption in structural engineering does not just underestimate a probability. It hides a phase transition: a sudden change of kind, like water turning to ice, rather than a gradual change of degree.

The Governing Formula

Structural inspection models assess building safety by evaluating the condition of individual elements and computing a composite score:

$$P(\text{structural failure}) \;\leq\; \sum_{i=1}^{n} p_i \tag{7.1}$$

where p_i is the failure probability of element i. This is the union bound: the probability that at least one element fails is at most the sum of individual failure probabilities. When probabilities are small, the approximation $P \approx \sum p_i$ is standard.

The union bound is tight when failures are independent. When failures are correlated—when elements degrade together—the probability of *simultaneous multi-element failure* is higher than the product of individual probabilities, and the failure mode changes from one element failing (which may be survivable through structural redundancy) to many elements failing at once (which causes progressive collapse).

The critical quantity is not $\sum p_i$ but the joint failure probability:

$$P(\text{multiple elements fail simultaneously}) \;=\; \prod p_i \quad \text{(independent)}. \tag{7.2}$$

The Independence Assumption

Structural elements in the same building are not independent. They share:

- **Environment.** Every element in Champlain Towers South was exposed to the same salt air, the same humidity, the same water table, and the same intrusion of chloride, the salt that corrodes steel. Reinforcing steel corrodes when chlorides penetrate the concrete cover. In a beachfront building, every element is

exposed to the same chloride source.

- **Materials.** Elements cast from the same concrete batch share the same mix design, the same aggregate, the same water-cement ratio, and the same curing conditions. A deficiency in the concrete—low cover depth, high porosity, inadequate consolidation—affects every element poured from that batch.

- **Construction practice.** The same construction crew, working under the same schedule pressure, makes the same errors across the building. Insufficient rebar cover, improperly tied connections, and inadequate waterproofing at expansion joints are systematic, not random.

- **Load redistribution.** When one element weakens, adjacent elements carry more load. The overloaded elements degrade faster. Degradation creates more load redistribution, which creates more degradation. The feedback loop couples the elements.

The Corrected Formula

For n structural elements sharing degradation drivers, with coupling d:

$$\boxed{P(m \text{ elements fail simultaneously}) \;=\; \binom{n}{m}(d \cdot p)^m (1 - d \cdot p)^{n-m}} \tag{7.3}$$

Here $\binom{n}{m}$ counts the ways of choosing which m of the n elements are the ones that fail.

The progressive-collapse risk—the probability that enough elements fail simultaneously to trigger cascading failure—depends on a threshold m^*: the minimum number of simultaneous element failures that exceeds the structure's redundancy.

Under independence ($d = 1$), $P(m \geq m^*)$ may be negligible. Under coupling ($d > 1$), it can be orders of magnitude higher.

Here is the phase transition: in the other domains of this book, coupling makes the bad outcome more probable. The supply chain is more likely to be disrupted; the grid is more likely to black out. But the *kind* of outcome is the same—a larger version of the failure the model already considers.

In a building, coupling changes the kind of failure. Below m^*, a structural failure is a crack, a spall, a repair bill. Above m^*, it is a progressive collapse: each failing element transfers its load to neighbors, which fail, which transfer more load, in a cascade that runs faster than any human can react to. The corrected formula does not just raise the probability of a known outcome. It reveals that above a threshold, the outcome itself transforms from repairable to catastrophic. The coupling is not just a multiplier. It is a trigger.

For Champlain Towers South:

The 2018 Morabito report found that the pool-deck waterproofing layer was not sloped, that the concrete slabs below had been severely damaged by the water it trapped, and that the damage would "expand exponentially" if the waterproofing was not replaced. NIST's technical investigation (June 2026) concluded that the collapse began when two connections between garage columns and the pool deck failed, and traced the root causes to problems in the original design and construction of the building. Whatever the mechanism, the elements at risk shared a design, a construction crew, and a water pathway. They were coupled.

For illustration, take 30 elements sharing these drivers. If the per-element annual failure probability were $p = 0.001$ and the coupling among them were $d \approx 3$ (reflecting shared design, materials, water exposure, and load redistribution)—both values are assumptions, not NIST estimates—then:

P(5 columns fail simultaneously):

Independent: $\binom{30}{5}(0.001)^5 \approx 1.4 \times 10^{-10}$.

Coupled: $\binom{30}{5}(0.003)^5 \approx 3.5 \times 10^{-8}$—a factor of $3^5 = 243$ higher.

The absolute probability is still small per year, but the building stood for 40 years, cumulating risk. And the per-column failure probability was not static: it rose with each year of corrosion, and the coupling rose with it as load redistributed from weakened columns to their neighbors.

The Counterfactual

The 2018 inspection of Champlain Towers South identified "abundant cracking and spalling" (concrete flaking away in chunks) in the parking garage columns and "major structural damage" to the concrete slab below the pool deck. The report warned that repairs would be "extremely expensive" and that delay would let the deterioration "expand exponentially." By an April 2021 letter to residents the remedial program was estimated at $15 million, and the work had not begun when the building collapsed.

The corrected inspection model would have changed the urgency assessment:

1. **Correlated degradation scoring.** The standard inspection evaluates each element independently: this column has moderate corrosion, that beam has light spalling. The corrected model scores the *coupling* between elements: are the degraded elements sharing a common cause (water pathway, concrete batch, load path)? If yes, the corrected model flags the group as a coupled cluster with $d > 1$, and the failure probability of the cluster is d^m times higher than the product of individual probabilities.

2. **Shorter re-inspection intervals for coupled elements.** The standard 40-year recertification interval is calibrated to independent degradation. The corrected interval for a building with high coupling is shorter: if $d = 3$ and $m = 5$ coupled elements,

the effective degradation rate is 3× the individual rate, and the inspection interval should be shortened by at least a factor of 3.

3. **Progressive-collapse assessment.** The current standard does not require an assessment of how many elements must fail simultaneously to trigger progressive collapse. The corrected standard requires computing m^* (the collapse threshold) and comparing it to the coupled failure probability. For Champlain Towers South, the NIST investigation found that the collapse began when two connections between garage columns and the pool deck failed. With $m^* = 2$, the joint-failure term applies directly: two coupled connections failing together is the event whose probability the coupling raises by a factor of d^2, and the shared design, construction, and water exposure are what make such a joint failure more likely than the product of the individual probabilities.

The Barrier

- **Inspection scope.** Structural inspections are visual and localized. An inspector examines accessible elements—typically in the parking garage and common areas—and rates their condition. Computing d requires a systematic survey of all elements in a potential failure group, with data on the shared drivers (water pathways, material batches, construction records). This is more labor-intensive and more expensive than the current practice.

- **Building codes.** Florida law (SB 4-D, 2022) now requires milestone structural inspections at 30 years, or 25 years within 3 miles of the coast, and every 10 years thereafter; a 2023 amendment lets local governments choose 25 or 30 years for coastal buildings. The amendment addresses the *frequency* of inspection but not the *model*: the inspections still evaluate

elements individually, not as coupled groups.

- **Liability.** An engineer who reports that individual elements are "satisfactory" faces less liability than one who reports that a coupled group of elements is "at risk of progressive collapse." The standard does not require the latter assessment, and the profession has not adopted it. The corrected model asks engineers to make a claim about joint failure probability, which is harder to defend than a claim about individual condition.

- **Condo association governance.** At Champlain Towers South, the warning that repairs would be "extremely expensive" was known for three years before the collapse. The condo association debated the cost, delayed the work, and sought lower bids. The corrected model would have flagged the urgency—but the decision to fund repairs is made by residents, not by engineers, and residents respond to perceived risk, not to a formula.

- **A second coupling.** There is a bitter irony in this domain. The structural elements are coupled through shared environment and materials. The decision-makers are coupled through shared financial exposure: every condo owner faces the same special assessment, reads the same engineer's report, and votes in the same meeting. If one owner votes against the repair, the information propagates: "maybe it's not that urgent." The coupling in the governance mirrors the coupling in the concrete. The system that needs to decide on the repair is itself a coupled system, with its own $d > 1$—and its own tendency to underestimate correlated risk.

What to remember. Structural inspection assumes elements degrade independently. The corrected model replaces p with dp for elements sharing environment, materials, and load paths: at $d = 3$, five coupled columns are 243× more likely to fail simultaneously. The correction calls for coupled-degradation scoring and shorter inspection intervals for buildings with high environmental coupling.

Chapter 8

The Redundancy

The question: What is the corrected safety-case formula for redundant systems, and when does adding a backup make the system less safe?

The word "independent" does more work than any other in the system safety assessment for a modern fly-by-wire aircraft, one whose controls run through computers rather than direct mechanical links. It appears throughout the documents engineers write to show that a system is safe: in the fault-tree analysis, in the failure-mode-and-effects analysis, in the common-cause analysis, and in the certification summary. It is the load-bearing word in every safety argument that multiplies probabilities.

When a certification engineer writes "the probability of simultaneous failure of all three hydraulic channels is $10^{-3} \times 10^{-3} \times 10^{-3} = 10^{-9}$," the word doing the work is not "probability." It is the implicit "and" between the three terms—the multiplication sign that assumes the channels fail for independent reasons.

The safety case for a flight-critical system begins with a number: 10^{-9}. One catastrophic failure per billion flight hours. This is the design target set by the Federal Aviation Administration (FAA) in Advisory Circular AC 25.1309 and its equivalents worldwide.

To meet this target, aircraft systems use redundancy: two or three independent channels, any one of which can perform the function. If a single channel has a failure rate of 10^{-3} per flight hour, two independent channels give 10^{-6}, and three give 10^{-9}. The arithmetic

is the multiplication rule for independent events:

$$P(\text{system failure}) \;=\; p^n \tag{8.1}$$

where p is the per-channel failure rate and n is the number of independent channels.

The Independence Assumption

"Independent" channels share:

- **Software.** The same flight-control software runs on all three channels of a triplex (three-copy) fly-by-wire system. A software defect—a logic error, a numeric overflow, a mishandled edge case—affects all channels simultaneously. The probability of a software-caused failure is not p^3; it is p, because the defect is in the specification, not in the hardware.

- **Design.** "Dissimilar redundancy" uses different hardware implementations of the same function. But the implementations share the same requirements document, the same operational concept, and often the same design team. A requirements error propagates to all implementations.

- **Environment.** All channels are in the same aircraft, exposed to the same temperature, altitude, vibration, electromagnetic environment, and bird strikes. An environmental event—a lightning strike, a volcanic ash encounter, a total loss of electrical power—affects all channels.

- **Maintenance.** A maintenance error—an incorrectly torqued connector, a wrong-revision part, a skipped inspection—may affect all channels that were serviced in the same maintenance visit.

- **Common-mode failures.** The aviation safety literature has recognized common-mode failures (CMFs) since the 1970s. The standard methodology (ARP4761, SAE International) includes a "common-cause analysis" step. But the analysis is qualitative: it lists potential common causes and argues, by engineering judgment, that each has been mitigated. It does not compute d.

The Corrected Formula

For an n-channel redundant system with common-mode coupling d:

$$\boxed{P(\text{system failure}) \;=\; (d \cdot p)^n} \tag{8.2}$$

Triplex redundancy with $p = 10^{-3}$ and $d = 10$ (a single shared software or design defect dominates):

$$P = (10 \times 10^{-3})^3 = (10^{-2})^3 = 10^{-6}.$$

The safety case claims 10^{-9}. The corrected formula gives 10^{-6}—a thousand times worse. The system still meets the "remote" threshold that applies to a major failure (10^{-5}) but misses the "extremely improbable" threshold that applies to a catastrophic failure (10^{-9}) by three orders of magnitude.

The extreme case: if the coupling channel is a software defect that is present in all channels ($d = 1/p$, meaning the coupled failure probability is 1), then $P = 1$ regardless of n. Adding channels does not help. The system's reliability is bounded by its least-independent failure mode.

The β-Factor Model

This section is for readers who work in reliability engineering; the algebra can be skipped. The conclusion is in the paragraph after the second equation:

the β-factor model is a special case of the d^n framework, one that needs expert judgment where d^n needs data.

The aviation and nuclear-safety communities have a partial solution: the β-factor model, which assigns a fraction β of each channel's failures to common causes: failures that take out all the channels at once. The total failure rate becomes:

$$P = (1-\beta)^n p^n + \beta\, p$$

where the first term is the independent fraction and the second is the common-cause fraction (which does not benefit from redundancy).

This model has been in use since the 1980s. It is a special case of the d^n framework. Setting the β-factor total equal to $(dp)^n$ and solving for d gives $d^n = (1-\beta)^n + \beta\, p^{-(n-1)}$. When the common-cause term dominates (i.e., $\beta\, p^{-(n-1)} \gg 1$, which is the regime that matters), the effective coupling is:

$$d \;\approx\; \beta^{1/n}\, p^{-(n-1)/n},$$

which grows sharply with n and can be far larger than 1 even for small β. At $\beta = 0.01$, $n = 3$, $p = 10^{-3}$: $d \approx 21.5$—a triplex system whose common-cause coupling dwarfs its independent-failure rate. In other words, a 1% common-cause share makes the three-channel system behave as if each channel failed 21.5 times as often. The β-factor model captures the same physics but in a less transparent form: it does not make explicit that the common-cause component grows as d^n with the number of redundant channels, and it requires estimating β from expert judgment rather than measuring d from data.

The d^n framework offers three advantages over the β-factor model:

1. d **is measurable.** The permutation test computes d from operational data: a matrix of flight hours by system channel, with failure indicators. This grounds the common-cause analysis in

data rather than judgment.

2. **The exponent is explicit.** d^n makes clear that adding channels to a coupled system increases the gap between the claimed safety level and the real one. The β-factor model obscures this because β is a flat fraction, not an exponential.

3. **The crossover is visible.** Once $d > p^{-(n-1)/n}$, the real failure probability $(dp)^n$ is larger than the single-channel rate p that the safety case started from: however many backups the safety case counts, the redundant system is no safer than one uncoupled channel. For $p = 10^{-3}$ and $n = 3$ that threshold is $d = 100$. The β-factor model does not surface this crossover.

The Counterfactual

The Boeing 737 MAX's MCAS system, which pushes the nose down to prevent a stall, used a single angle-of-attack (AoA) sensor—a probe that measures the angle between the wing and the oncoming air—as input, with no cross-check against the second AoA sensor. When the active sensor gave erroneous data (in the Lion Air case, from a sensor that had been replaced two days before the accident flight), MCAS commanded repeated nose-down trim that the crew could not overcome. 346 people died.

This was not a d^n error in the classical sense—the system did not have redundant AoA inputs. It was a design choice that violated the independence assumption at a more basic level: the flight-control function depended on a single sensor despite the existence of a second sensor on the airframe. But the broader pattern applies: the safety case for the 737 MAX relied on the crew as a redundant backup to the automation. The crew and the automation shared a coupling channel—the same confusing cockpit indications, the same time pressure, the same flight-control response—with $d \gg 1$.

The corrected safety case, applied to human–automation re-

dundancy, would have flagged the crew-as-backup assumption as coupled: a crew presented with a confusing automation failure is not an independent backup but a coupled one, with d estimated from simulator studies of crew response to automation failures. No such estimate is established here; the point is that the safety case should have to supply one.

The Barrier

- **Certification path dependency.** The 737 MAX was certified as a derivative of the 737 NG, which was a derivative of the 737 Classic, which descended from the original 737 certified in 1967. Each derivative was required to demonstrate equivalence to the previous model, not to meet a new safety standard from scratch. The corrected formula would require re-assessing common-mode coupling for each derivative—a process that the derivative certification pathway is designed to avoid.

- **The independence claim is the safety case.** The entire architecture of aviation safety rests on the claim that redundant channels are independent. Acknowledging $d > 1$ does not just change a number; it changes the validity of thousands of safety cases for every system on every aircraft in service. The incentive to maintain the independence assumption is proportional to the cost of reopening those cases.

- **Qualitative vs. quantitative common-cause analysis.** ARP4761 requires a common-cause analysis but allows it to be qualitative. The corrected framework requires a quantitative d, which requires operational failure data from redundant channels. This data exists in airline maintenance records and in service-difficulty reports, but it has not been systematically analyzed for common-mode coupling.

- **The rare-event problem.** Catastrophic failures in aviation are extremely rare—which is the success of the current system. The rarity means that d cannot be measured from catastrophic-failure data directly. It must be estimated from precursor events (near-misses, in-flight shutdowns, maintenance findings), which are more common but less directly relevant to the catastrophic case.

What to remember. The safety case for redundant systems assumes independent channel failures. The corrected formula replaces p with dp: triplex redundancy with $d = 10$ gives 10^{-6}, not 10^{-9}—a thousand times worse than claimed. The correction calls for data-driven common-mode coupling measurement, replacing qualitative common-cause analysis with quantitative d.

Chapter 9

The Ensemble

The question: What is the corrected ensemble uncertainty, and how many independent models does an ensemble of forty actually contain?

This chapter is unusual in this book, because the corrected formula does not reveal a hidden danger. It reveals a hidden ignorance.

In the other six domains, the correction shows that the risk is higher than the model claims: more failures, more deaths, more dollars. In climate science, the correction shows that the uncertainty is wider than the model claims—in both directions. The correction says we might warm more than the headline numbers suggest, *and* we might warm less. This makes the correction politically inconvenient for everyone: for activists who cite the narrow range as settled certainty, and for skeptics who cite it as alarmist precision. Honest uncertainty is a hard sell.

The Intergovernmental Panel on Climate Change (IPCC) bases its projections on multi-model ensembles. The Coupled Model Intercomparison Project, now in its sixth phase (CMIP6), includes output from dozens of global climate models developed by research groups worldwide (about 40 in the sensitivity analyses discussed here). The models are run under shared forcing scenarios, and their outputs are compared to produce a range of projected futures.

The range is interpreted as uncertainty. The spread of the models is treated as a proxy for how much we do not know.

The Governing Formula

The standard error of the ensemble mean is:

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}} \tag{9.1}$$

where σ is the standard deviation of the model outputs and n is the number of models. With $n = 40$, the standard error is $\sigma/\sqrt{40} \approx 0.158\,\sigma$. The standard error measures how uncertain the average of the models still is: the smaller it is, the more confident we can be in that average.

This formula assumes that each model is an independent draw from the space of possible climate models. The uncertainty decreases as $1/\sqrt{n}$ because each new model adds independent information.

The Independence Assumption

Imagine asking forty people to estimate the height of a building. If each looks at it alone, their average is far more reliable than any single guess. If thirty-five of them copied from one another, you do not have forty opinions; you have a handful. A climate-model ensemble is closer to the second case.

The models are not independent draws. They share:

- **Code heritage.** Many CMIP6 models descend from a small number of ancestral codes. CESM, GFDL, and UKESM share components (ocean models, land-surface models, radiation schemes) that evolved from common ancestors. A structural choice made in an ancestral code—a parameterization of cloud microphysics, a convective scheme, a treatment of sea ice—propagates to all descendants.

- **Parameterization choices.** The models share a scientific culture. When a parameterization paper is widely cited, multiple

modeling groups adopt similar approaches. The "diversity" of parameterizations is narrower than the diversity of possible approaches because the groups read the same literature and attend the same conferences.

- **Observational constraints.** All models are calibrated against the same observational record—the same sea-surface temperature data, the same reanalysis products (observations blended into one gridded record), the same satellite retrievals. A bias in the observational record (e.g., sparse coverage in the Southern Ocean before satellite era) biases all models in the same direction.

- **Tuning targets.** Modeling groups tune their models to reproduce key features of the current climate: the global mean temperature, the strength of the Atlantic overturning circulation, the El Niño period. The tuning targets are shared, and meeting them constrains the models to agree—reducing spread without adding independent information.

The Corrected Formula

The corrected standard error replaces n with the effective sample size n_{eff}:

$$\boxed{\text{SE}_{\text{corrected}} = \frac{\sigma}{\sqrt{n_{\text{eff}}}} = \frac{\sigma}{\sqrt{n/(1+(n-1)\bar{\rho})}}} \tag{9.2}$$

where $\bar{\rho}$ is the average pairwise correlation between model outputs. In words, n_{eff} is the number of truly independent models that would carry as much information as your n correlated ones. When $\bar{\rho} = 0$ (independent), $n_{\text{eff}} = n$, recovering the standard formula. When $\bar{\rho} > 0$, $n_{\text{eff}} < n$, and the uncertainty is wider than the standard formula claims.

The connection to d: the permutation test applied to the model-output matrix (rows are years or grid cells, columns are models) gives a pricing ratio $R = d^n$ for the co-occurrence of model agreement on extreme outcomes. The effective sample size is related to the pairwise correlation, which is in turn related to d through the dependence structure. For practical purposes, the pairwise correlation $\bar{\rho}$ is the more directly useful quantity for ensemble uncertainty; the d^n framework provides the theoretical foundation.

Illustrative inter-model correlations (assumed values; published estimates depend on the variable and the method):

Variable	$\bar{\rho}$ (approx.)	n_{eff} from 40 models
Global mean temperature trend	0.20–0.25	3.7–4.5
Regional precipitation	0.12–0.20	4.5–7.0
Arctic sea ice extent	0.25–0.30	3.2–3.7
Equilibrium climate sensitivity	0.15–0.20	4.5–5.8

With these illustrative correlations, an ensemble of 40 models behaves like 3 to 7 independent ones; published analyses of CMIP model dependence (Knutti et al., Sanderson et al.) likewise find effective numbers of a handful, far below the model count.

The Consequence

The standard uncertainty estimate from 40 models:

$$\text{SE}_{\text{standard}} = \frac{\sigma}{\sqrt{40}} \approx 0.158\,\sigma.$$

The corrected estimate at $n_{\text{eff}} = 4$:

$$\text{SE}_{\text{corrected}} = \frac{\sigma}{\sqrt{4}} = 0.5\,\sigma.$$

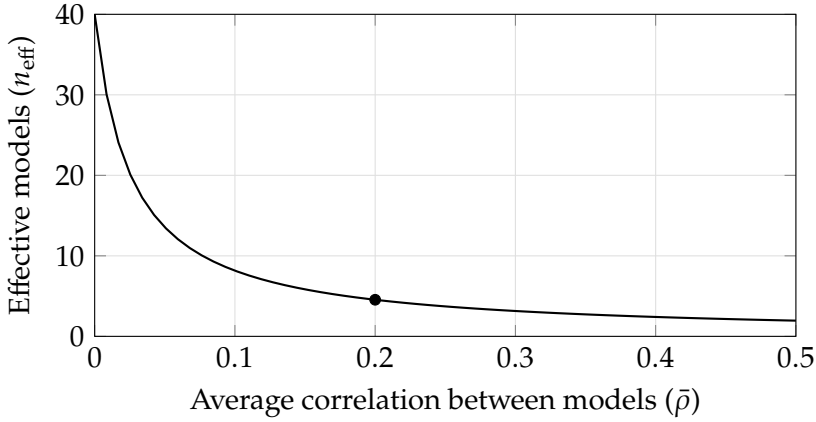

Figure 9.1: How many independent models an ensemble of forty is worth, as the average correlation between its members rises. The curve is $n_{\text{eff}} = 40/(1 + 39\bar{\rho})$. The dot marks $\bar{\rho} = 0.2$, where forty models carry the information of about four and a half.

The corrected confidence interval is $0.5/0.158 \approx 3.2$ times wider than the standard-error interval.

This does not mean that climate projections are wrong. It means that the uncertainty attached to a multi-model ensemble mean is wider than the independent-model formula reports. For an interval built as ± 1.645 standard errors from a model spread σ, the standard formula gives $\pm 0.26\,\sigma$ and the corrected one $\pm 0.82\,\sigma$. The central estimate is unchanged; the interval is about three times as wide.

The Counterfactual

The corrected ensemble uncertainty would not have changed the central estimate of warming. It would have changed how far a multi-model ensemble can be trusted to narrow the range of futures.

The IPCC Sixth Assessment Report (2021) is a useful case, and a careful one. It reported a "likely" range for equilibrium climate sensitivity (the long-run warming from a doubling of atmospheric

carbon dioxide) of 2.5 to 4.0°C, a significant narrowing from the Fifth Assessment's likely range of 1.5 to 4.5°C. It did not get there by averaging models. It combined several lines of evidence—process understanding of feedbacks, the historical record, and paleoclimate—and gave less weight to the highest-sensitivity CMIP6 models, whose raw range ran from 1.8 to 5.6°C. The correction in this chapter therefore does not change that range. It applies to statements that do rest on ensemble statistics: multi-model mean projections of temperature, precipitation, or sea level, and the confidence intervals attached to them.

For those statements, the corrected formula would have:

1. Widened the stated uncertainty of ensemble-mean projections to reflect 3–7 effective models rather than 40.
2. Kept attention on the high-end outcomes that the ensemble spread allows, instead of letting a narrow standard error suggest they were excluded.
3. Honestly represented the state of knowledge: this ensemble offers 3–7 effective independent estimates, not 40.

The Barrier

- **Political sensitivity.** Wider uncertainty ranges are perceived as weakening the case for climate action. "We are less sure" is harder to translate into policy than "we are likely to see 2.5 to 4.0°C." The correction says something that is scientifically honest and politically inconvenient: we are less certain than the headline numbers suggest, in both directions.
- **The independence fiction is useful.** Treating 40 models as 40 independent estimates produces impressively narrow confidence intervals. The narrowness conveys scientific authority. Acknowledging $n_{\text{eff}} = 4$ would force a reckoning with how much

of the apparent precision is an artifact of counting dependent models as independent.

- **Model weighting is contested.** Several research groups have proposed model-weighting schemes that account for dependence (Knutti et al., 2017; Brunner et al., 2020). These methods are technically sound but require subjective choices about how to define independence, and different weighting schemes give different n_{eff}. The field has not converged on a standard approach.

- **Model development incentives.** Modeling centers are funded to develop and contribute models to CMIP. A framework that says "your model is not independent of the five others that share its ocean component" reduces the scientific credit for contributing a model. The incentive structure rewards model production, not model independence.

What to remember. Ensemble uncertainty assumes independent models. Forty CMIP6 models behave like a handful of independent ones (3 to 7 with illustrative correlations). The corrected confidence interval is roughly 3× wider than the independent-model formula gives. The correction does not change the central estimate; it widens the tails of ensemble-based projections.

Seven Formulas at a Glance

The seven chapters of Part II follow one pattern: a formula that the field trusts, the shared cause it ignores, and what happens when the shared cause is put back. The table collects them. The effects are the worked examples from each chapter; the values of d are the ones used there and, as each chapter says, are illustrative except where stated.

Domain	Formula trusted	Cause it ignores	Corrected effect (worked example)
Supply chains	$1 - q^n$	Same foundry, strait, and just-in-time buffer	Six suppliers, $d = 1.8$: total failure 34× likelier
Software	$1 - f^n$	Same kernel driver, cloud, and update timing	Two backups, $d = 5$: six nines becomes four and a half
Power grids	$\prod q_i$	Same weather and gas pipeline	$d = 2.5$: five outages 97× likelier
Medicine	$1 - \prod(1 - r_i)$	Same liver enzyme	$d = 2$, five drugs: 9.6% becomes 18.5%
Buildings	$\binom{n}{m} p^m$	Same salt, concrete, and load path	$d = 3$: five columns failing together 243× likelier
Aviation	p^n	Same software, design, and environment	Triplex, $d = 10$: 10^{-9} becomes 10^{-6}
Climate	$\sigma/\sqrt{n}$	Same code, tuning, and data	Forty models act like 3 to 7; interval about 3× wider

Part III turns from correcting formulas to keeping the correction

current: how to monitor d, how it is changing, and why the corrected formulas have not yet been adopted.

Part III

The Practice

Chapter 10

The Monitor

The question: How do you track coupling in real time, and what do you do when it spikes?

Imagine a dashboard. Not the kind that shows whether each component is up or down—every organization has that. A different kind: one number, updated weekly, that tells you how wrong your model is right now.

The number is d. When it is near 1, your model is approximately correct. When it is 2, your model is wrong by 2^n, where n is the number of components that must fail together. When it reaches 3, the model is wrong by thousands at $n = 7$. The number does not tell you *what* will fail. It tells you that when something fails, it will be worse than your model predicts, and by how much.

The corrected formulas in the preceding chapters share a feature: they all contain d. And d is not a constant. It changes with time, with conditions, and with the system's own evolution.

Supply-chain d was low in 2019 and high in 2021. Grid d is low in mild weather and high during extreme events. Software d is low during normal operation and spikes during vendor incidents. The corrected formula is only as good as the d you put into it.

This chapter describes a monitoring system: how to compute d from rolling data, how to display it, what levels trigger action, and what the action is.

Your First Measurement

Before monitoring anything over time, measure once. There are five steps.

1. **Name the event.** Write down the compound event your formula prices: at least M of your components in distress in the same period, and, if it matters, for K periods in a row.

2. **Build the matrix.** Rows are periods, columns are components, and each entry is 1 if that component was in distress in that period and 0 otherwise. Decide what "distress" means before you look at the answer.

3. **Count the event in the real data.**

4. **Count it in shuffled data.** Shuffle each column independently, count the event again, and average over many shuffles.

5. **Divide, then take the root.** R is the real count divided by the shuffled count. Then $d = R^{1/n}$, where n counts the "ands" in the event: one for each component that must be in distress, plus one for each further period it must last.

The same steps as a sketch:

```
X     = 0/1 matrix (rows: periods,
        columns: components)
event = at least M columns equal 1 in
        the same row (for a streak:
        in K consecutive rows)
real  = times the event occurs in X
null  = average times it occurs over
        B copies of X, each with every
        column shuffled independently
R     = real / null
d     = R ** (1 / n)
        # n = number of "ands"
```

Three cautions. If the shuffled count is zero, R is unbounded: use more shuffles or a less demanding event, and report d as a lower bound. If the event appears in the real data only a handful of times, treat d as a rough guide. And if the data has a strong seasonal cycle, shuffle within each season (for daily data, within each calendar month), or the shared seasons will be counted as coupling.

The Rolling Measurement

The permutation test from *The Rogue Wave* computes R (and hence d) for a static matrix. To monitor d over time, run the test in a rolling window:

1. Choose a window length. One year of weekly data ($n_t = 52$ rows) is typical; shorter windows are noisier, longer windows are smoother but slower to respond.

2. At each time step, construct the matrix from the most recent window of data.

3. Run the column shuffle (B is the number of random shuffles: $B = 200$ is sufficient for monitoring; $B = 1{,}000$ for publication-quality estimates).

4. Record the pricing ratio R and compute $d = R^{1/n}$ where n counts the "ands" in the cost function (for a cost that requires M components in distress at once, $n = M$).

5. Advance the window by one step and repeat.

The output is a time series of d.

Four Panels

A monitoring dashboard displays four quantities:

1. **Current d.** The most recent rolling-window measurement. This is the number that goes into the corrected formula.

2. **Lag structure.** The lag-shift test from *The Rogue Wave* measures whether coupling is instantaneous (components move together on the same day) or delayed (component A leads component B by days or weeks). Instantaneous coupling (lag = 0) means no warning window: by the time you detect the coupling, the compound event is already happening. Delayed coupling means you may have time to act.

3. **Co-movement ratio $R(M)$.** The full pricing ratio for M or more components in simultaneous distress. This shows the operational severity: at $M = 3$ out of 6, what is the model error? At $M = 5$ out of 6?

4. **Trend.** The 5-year or 10-year moving average of d. This shows whether the system is becoming more or less coupled over time. A rising trend means the corrected formula must be recalibrated—and it means the risk is growing even if no event has occurred.

Response Levels

The following levels are illustrative thresholds, not fitted values. They are approximate—the domains differ, and the appropriate thresholds depend on the system's tolerance for model error.

Level	d range	Response
Normal	$d < 1.5$	Standard operations. The model error is at most 1.5^n; for $n = 5$, that is 7.6×. Non-trivial but within the safety margins of most systems.
Elevated	$1.5 \leq d < 2.5$	Increase safety margins by d^n. Extend inventory buffers. Increase reserve capacity. Widen ensemble confidence intervals. The model is wrong by 7.6× to 97.7× at $n = 5$.
Severe	$d \geq 2.5$	The model is wrong by more than 97× at $n = 5$. Activate contingency plans. Reduce exposure to coupled components. At $d = 3$ and $n = 10$: the model is wrong by 59,000×. The formula without the correction is not a useful guide.

These thresholds are judgment calls: they are chosen so that the model error at $n = 5$ crosses roughly 8× and 100×. Different systems will want different cut-offs, and a system's d can move substantially between calm and crisis.

What Changes When d Moves

The corrected formula is a function of d. When d doubles, the model error at $n = 5$ increases by $2^5 = 32\times$. This means:

- **Inventory buffers** (supply chains) should scale with d^n, not with a fixed safety stock. A procurement system that monitors d and adjusts buffer levels dynamically—larger stocks when d is elevated, normal stocks when d is low—would have been pre-positioned for the 2021 shortage.
- **Reserve margins** (power grids) should increase during weather

events. The grid operator who monitors d between generators in real time can call for additional reserves when d exceeds the elevated threshold—before the outages begin.

- **Staged rollouts** (software) should slow down when d between customers is high. A vendor that monitors co-failure rates across its customer base can detect abnormal coupling (indicating a systemic issue) and halt deployment before the blast radius expands.
- **Prescribing alerts** (medicine) should trigger when a patient's medication list crosses the coupling threshold for a shared metabolic pathway. An electronic health record system that computes d for the patient's drug combination and flags elevated coupling is technically feasible today.
- **Inspection frequency** (construction) should increase when environmental monitoring (chloride levels, water intrusion rates) indicates elevated coupling between structural elements.

The Data Requirement

Monitoring requires data. The minimum is a matrix: rows are time periods, columns are components, entries are distress indicators.

Not every domain has this data in a convenient form:

Domain	Data availability	Latency
Finance	Excellent (real-time prices)	Seconds
Software	Excellent (telemetry)	Minutes
Power grids	Good (SCADA, PMU)	Minutes
Supply chains	Moderate (AIS, PMI)	Days to weeks
Medicine	Moderate (EHR, FAERS)	Weeks to months
Construction	Poor (inspections)	Years
Climate models	Good (CMIP archives)	Months

(SCADA and PMU are the sensor feeds that report grid conditions second by second; AIS is ship tracking; EHR is the electronic health record; FAERS is the FDA's adverse-event database.)

Domains with high data availability and low latency (finance, software, grids) can monitor d in near real time. Domains with low availability and high latency (construction, medicine) can compute d periodically—annually or at each inspection—and use the trend to adjust safety margins.

The correction does not require real-time monitoring to be useful. Even a single measurement of d, computed once from historical data and plugged into the corrected formula, is better than the current practice of assuming $d = 1$.

What to remember. d is not a constant. It changes with conditions, and the corrected formula is only as good as the d you use. A monitoring system computes d in a rolling window, displays four quantities (current d, lag, co-movement ratio, trend), and triggers graduated responses: standard, elevated, severe. Even a single measurement of d is better than assuming $d = 1$.

Chapter 11

The Trend

The question: Is coupling getting stronger over time, and what does that mean for the formulas?

In 1995, the global semiconductor supply chain spread its fabrication across many countries. By 2021 a single company—TSMC—produced more than half of the world's foundry revenue, and advanced fabrication had concentrated in a handful of places. The supply chain did not become more coupled because of a disaster. It became more coupled because concentration is cheaper.

The corrected formulas in this book use a value of d. But if d is rising, the formula corrected with today's d will be wrong tomorrow.

Four Domains Where Coupling Is Rising

Supply chains are more concentrated than they were two decades ago. In 2000, semiconductor fabrication was distributed across the United States, Japan, South Korea, Taiwan, and Europe. By 2021, TSMC alone accounted for more than half of global foundry revenue. Geographic concentration is a coupling channel. The trend toward concentration was driven by economies of scale: each generation of fabrication technology requires larger investment, and fewer firms can afford it. The CHIPS Act (2022) is an attempt to reverse this trend for one sector; it does not address concentration in raw materials, chemical supply, or equipment manufacturing.

Software infrastructure is more consolidated. In 2010, cloud computing was nascent and most organizations ran their own data centers. By 2024, three providers (AWS, Azure, GCP) hosted the majority of enterprise workloads. The CrowdStrike event was possible only because a single vendor had achieved sufficient market penetration to make 8.5 million simultaneous failures a realistic scenario. The trend is accelerating: cloud migration continues, SaaS adoption grows, and the number of critical single vendors in an organization's dependency graph increases.

Power grids are more interconnected. The push for renewable energy requires long-distance transmission to move power from windy and sunny regions to demand centers. Interconnection increases the size of the coupled system: a weather event in one region can propagate through transmission to affect distant regions that would previously have been isolated. At the same time, the transition from dispatchable generation (gas, coal, nuclear) to weather-dependent generation (wind, solar) increases the coupling between the grid and the weather.

Medicine has more polypharmacy. The percentage of US adults taking five or more prescription medications rose from 8.2% in 1999–2000 to 15% in 2011–2012, and it is higher among older adults. Each additional drug on a shared metabolic pathway increases n in the d^n formula. The trend is demographic (aging populations take more medications) and medical (new drugs are developed for conditions that were previously untreated).

The Numbers

How much has coupling changed? Precise historical d measurements are unavailable—nobody was running the permutation test in 2000—but the structural drivers can be quantified:

Domain	Coupling driver	~2000	~
Semiconductors	TSMC foundry share	—	> 50
Cloud	Top-3 provider share	—	
Medicine	Adults on ≥5 prescription drugs	8.2% (1999–2000)	15% (2

These are proxies, not measurements of d. Only the medicine row has both endpoints measured; the others show a recent level and the direction the text describes.

A rough projection: if d in a domain rises from 1.5 in 2000 to 2.0 in 2024 (a 33% increase over 24 years, consistent with the concentration trends above), the pricing ratio at $n = 5$ rises from $1.5^5 = 7.6$ to $2.0^5 = 32$—a fourfold increase in model error. At the same linear rate, $d \approx 2.3$ by 2040, giving $2.33^5 \approx 69\times$. A model that was usable in 2000 (off by 8×, within a safety margin) becomes dangerous by 2040 (off by roughly 70×, outside any reasonable margin).

The Acceleration

If d grows linearly over time and n is constant, the pricing error d^n grows as a power of time; if d compounds at a constant percentage rate, it grows exponentially. If both d and n grow—more coupling per component *and* more components in the system—the error grows faster than any fixed exponential.

The short algebra below checks the last claim; it can be skipped.

Let $d(t) = d_0 + \alpha t$ and $n(t) = n_0 + \beta t$. The pricing ratio is:

$$R(t) = d(t)^{n(t)} = (d_0 + \alpha t)^{n_0 + \beta t}.$$

Taking logarithms:

$$\ln R(t) = (n_0 + \beta t) \ln(d_0 + \alpha t).$$

Both factors grow, so $\ln R$ grows faster than linearly—and R itself

grows faster than any fixed exponential. The exponent itself keeps growing: the model error does not just grow; it accelerates.

A small example shows the size of the effect. A system with $d = 2$ and $n = 5$ has $R = 2^5 = 32$. Let both grow modestly, to $d = 2.3$ and $n = 6$: then $R = 2.3^6 \approx 148$. A 15% rise in coupling and one more component made the error more than four times larger.

In supply chains, both are rising: more geographic concentration (d up) and more tiers of dependency (n up, as supply chains add intermediaries and subcontractors). In software, both are rising: more vendor consolidation (d up) and more services per application (n up, as microservice architectures proliferate). In medicine, both are rising: more drug interactions (d up from pathway crowding) and more drugs per patient (n up from polypharmacy).

The corrected formulas in this book are calibrated to current d and current n. If the trends continue, recalibration is not optional—it is necessary to avoid the corrected formula itself becoming stale.

The Window

There is a window between when the correction becomes necessary and when the next compound failure forces it. We are inside that window now.

In supply chains, the post-pandemic period of elevated d produced the $210 billion semiconductor shortage. The industry responded with geographic diversification (CHIPS Act, new fabs in Arizona and Germany). These investments lower one coupling channel (geography) while the others (shared equipment vendors, shared chemical suppliers, shared design tools) continue to concentrate. The net effect on d is uncertain. What is certain is that nobody is measuring it.

In software, the CrowdStrike event produced regulatory interest in vendor concentration. The EU's Digital Operational Resilience Act (DORA, effective January 2025) requires financial institutions to assess

ICT concentration risk. It does not specify how. The permutation test is one answer; it has not been proposed.

In medicine, polypharmacy is growing faster than the pharmacovigilance system can track. FAERS receives well over a million reports per year; estimates of the share of actual adverse events that are ever reported run from under 1% to about 10%. Most of the information needed to compute d is never recorded.

In power grids, FERC ordered a review of cold-weather reliability standards after Uri. The resulting standards, most recently EOP-012-2 (approved in June 2024), require winterization plans but does not require coupled outage-rate modeling. The coupling term is still missing from the standard.

The correction is more urgent today than it would have been twenty years ago. The independence assumption was always wrong, but it was less wrong when systems were less coupled and less complex. The formula that was off by 8× in 2000 is off by 32× in 2024 and will be off by roughly 70× in 2040, if the trends continue.

The corrected formula is not a one-time fix. It is a commitment to measuring d periodically and updating the formula when d changes.

What to remember. Coupling is rising across domains: supply-chain concentration, software consolidation, grid interconnection, polypharmacy. If d compounds at a constant rate and n is constant, the error grows exponentially. If both grow, the error grows faster than any fixed exponential. The correction must be recalibrated.

Chapter 12

The Barrier

The question: Why hasn't the correction been adopted, and what would it take?

The corrected formula is one line. Replacing p with dp takes three keystrokes. The permutation test that computes d is forty lines of code. A procurement analyst can run it on a laptop in ten minutes. A grid operator can run it on SCADA data in an afternoon. A pharmacist cannot run it at all, because the data is locked in separate EHR systems, but the computation itself is trivial.

The obstacle is not the math. The math has been available since the preceding chapters were written. The obstacle is that the formula lives inside a standard, the standard lives inside a regulation, and the regulation lives inside an institution.

The Rogue Wave devoted its final chapter to the independence trap: four forces—cognitive, institutional, selective, and methodological—that make the independence assumption self-repairing. That chapter explained why the error persists. This chapter asks what it would take to correct it.

The preceding chapters identify domain-specific barriers. This chapter identifies three barriers that cut across all domains. The table shows where each domain's barriers fall.

Domain	Standard	Data	Incentive
Supply chains	No standard requires coupling to be reported	Most firms cannot see past their tier-1 suppliers	Concentration is cheaper; the firm that builds buffers pays first
Software	SLAs are bilateral and say nothing of coupling between customers	Uptime is reported per service, not per dependency chain	Fast updates beat staged rollouts; concentrated markets
Power grids	N-1 and the NERC process are slow to change	Planning tools are built on independent outage rates	Who pays for the extra reserves
Medicine	Drugs are approved one at a time	Trials cannot test every combination; observational data ranks lower	Each specialist prescribes alone
Buildings	Codes set inspection frequency, not the model	Inspections are visual and local	Engineers' liability; owners' governance
Aviation	ARP4761 allows qualitative common-cause analysis	Catastrophic failures are too rare to measure d directly	Reopening thousands of safety cases is costly
Climate	No agreed way to define model independence	n_{eff} depends on the method chosen	Wider ranges are unwelcome; credit rewards model count

The First Barrier: The Standard

Every domain has a governing standard: NERC N-1 for grids, ARP4761 for aviation, Basel III for banking, the FDA approval framework for drugs, ISO 22301 for business continuity. The standard encodes the independence assumption. Correcting the formula means changing the standard.

Changing a standard is a different problem from proving it wrong.

Proof requires data and mathematics. Change requires consensus, transition planning, cost allocation, and political will. The data in this book proves the formulas wrong. Changing the standards requires something this book cannot provide: a coalition of practitioners, regulators, and policymakers who conclude that the cost of change is less than the cost of the next compound failure.

What this book can provide is the economic argument: the expected cost of the uncorrected formula, in each domain, exceeds the cost of correction. The preceding chapters make this argument case by case. The aggregate argument is simpler: across the seven domains in this book, the compound failures of the last decade—the semiconductor shortage ($210B), Winter Storm Uri ($195B), CrowdStrike ($5.4B), the opioid co-prescription crisis (tens of thousands of lives), Surfside (98 lives)—sum to hundreds of billions of dollars and thousands of deaths. The corrected formula, applied before each event, would have flagged the vulnerability at a cost of measurement (running the permutation test) and adjustment (changing safety margins, buffers, or inspection intervals). The measurement cost is negligible. The adjustment cost is real but small relative to the failure cost.

The Second Barrier: The Data

The corrected formula requires d, and d requires data. In some domains (finance, software, grids), the data exists and is abundant. In others (supply chains, medicine, construction), the data exists but is fragmented, proprietary, or not collected in a format suitable for the permutation test.

The data barrier is not a scientific problem. It is an infrastructure problem. The solutions are known:

- **Standardized disruption reporting.** Supply chains need a common format for reporting node-level disruptions, analogous

to the financial sector's trade-reporting requirements. The data exists in shipping records, customs databases, and procurement systems; it has not been aggregated.

- **Linked health records.** Medicine needs electronic health records that compute multi-drug coupling, not just list drug interactions. The data exists in EHR systems; the computation has not been implemented.

- **Sensor integration.** Construction needs structural health monitoring systems that correlate sensor data across elements, not just flag individual elements. The sensors exist; the correlation analysis has not been standardized.

The Third Barrier: The Incentive

The firm that corrects its formula unilaterally bears the cost (larger buffers, higher reserves, slower deployments) while its competitors do not. In a competitive market, the correcting firm is at a disadvantage every quarter that nothing goes wrong. The independent model is cheaper—until the compound event arrives.

This is a collective-action problem: everyone would be better off if all firms corrected, but each firm that corrects alone pays for it. The solution is regulatory: require the correction across the industry, so that no firm bears a competitive penalty for adopting it. The precedent exists. Capital requirements in banking (Basel III, the international banking rulebook) force all banks to hold reserves, preventing a race to the bottom. Safety standards in aviation (ARP4761) apply to all manufacturers. Environmental standards prevent firms from externalizing pollution costs.

The corrected formula is analogous. The uncorrected formula externalizes the cost of compound failure onto the public—the residents who freeze during Uri, the passengers stranded during CrowdStrike, the patients harmed by drug interactions, the residents of Champlain

Towers South. The corrected formula internalizes that cost, as all safety regulations do.

What Would Be Needed

For each domain, the minimum change:

Domain	Minimum change
Supply chains	Require d reporting in annual supply-chain risk assessments for critical industries.
Software	Include coupling metrics in SLA disclosures and business-continuity certifications.
Power grids	Amend NERC standard to require coupled outage-rate analysis during extreme-weather scenarios.
Medicine	Add multi-drug coupling assessment to EHR clinical decision support systems.
Construction	Require correlated-degradation analysis for buildings with shared environmental exposure.
Aviation	Amend ARP4761 to require quantitative common-mode coupling measurement from operational data.
Climate	Report n_{eff} alongside ensemble size in CMIP publications.

None of these changes requires new science. The test is published. The code is available. The data exists or can be collected. The barrier is institutional, not technical.

The Early Adopters

The correction has not been formally adopted by any standard-setting body. But the practice it implies—measuring coupling and adjusting for it—has been adopted, informally and under different names, by organizations that learned the hard way.

Toyota. After the 2011 Tōhoku earthquake disrupted its supply chain, Toyota built a supply-chain mapping system that tracks dependencies across multiple supplier tiers. It is widely reported to have held larger semiconductor stockpiles, which cushioned the early months of the 2021 shortage, although it too cut output sharply that September. Toyota does not use the word "coupling" or compute d. It does the equivalent: it maps the shared dependencies between its suppliers and maintains buffers proportional to the concentration risk. Toyota still posted a record operating profit of about 3 trillion yen for the fiscal year to March 2022, despite the shortage.

Netflix. After suffering outages from AWS dependency, Netflix built Chaos Monkey (2011) and its broader Simian Army—tools that randomly kill services in production to test resilience. The philosophy: if your system survives random component failure but not correlated failure, your redundancy is illusory. Netflix does not compute d, but its chaos engineering practice is an empirical version of the corrected formula: it tests for coupled failures, not just independent ones.

The Bank of England. In 2024 the Bank ran a System-Wide Exploratory Scenario (SWES) exercise to examine how the behavior of banks and non-bank financial institutions can amplify a shock, for example when one institution's distressed selling forces losses on others. Asking how much a shock is amplified by institutions acting on each other is asking how strongly they are coupled; that is the same question the correction answers, under another name.

These are not theoretical examples. They are organizations that paid the cost of a compound failure and changed their practice.

What they lack is the framework: a single measurement (d), a single correction ($p \to dp$), and a single test (the permutation test) that unifies their ad hoc responses into a quantitative standard. They are doing the correction by hand. The formula makes it systematic.

What the First Adoption Looks Like

The lowest-friction path is not regulatory. It is commercial.

A reinsurance company that computes d for its portfolio can price catastrophe bonds more accurately. It does not need regulatory approval to use the corrected formula; it needs better risk pricing. If the corrected formula prices a hurricane-season cat bond at $14 million instead of $9 million, the reinsurer captures the $5 million gap. The first mover profits.

A supply-chain risk analytics company that adds d to its dashboard gives its customers a number their competitors cannot see. The procurement officer who measures $d = 2.1$ across her supplier portfolio has information that the officer using the standard model does not. The correction is a competitive advantage before it is a regulatory requirement.

A hospital system that adds a coupling score to its EHR—flagging patients on 5+ drugs sharing a CYP pathway, with a d-derived risk estimate—reduces adverse events and malpractice exposure. The first health system to implement it reduces readmissions. The data to prove the reduction creates the evidence base for the regulation.

In each case, the pattern is the same: the correction creates private value before it creates public value. The early adopter does not bear a cost; it captures an information advantage. The regulation follows the proof of concept, not the other way around.

What to remember. Three barriers block adoption: the standard, the data, and the incentive. But early adopters—Toyota, Netflix, the Bank of England—already practice the correction under other names. The lowest-friction path is commercial: better risk pricing, better dashboards, better clinical alerts. The correction creates private value before it requires public mandate.

Epilogue

This book corrected seven formulas. There are hundreds more.

The Rogue Wave, Chapter 11, lists twelve domains where the independence assumption misprices compound events, and several more where it predicts the mispricing will be found. This book chose seven—the seven where the governing formula is most explicit, the data most available, the failure most vivid, and the correction most concrete.

The other domains are no less important. Insurance catastrophe models that assume regional independence. Epidemiological models that assume well-mixed populations. Agricultural risk models that assume independent regional yields. Genomic association studies that assume independent variant effects. Ecological models that assume additive species contributions. Nuclear safety assessments that assume independent barrier failures. Financial stress tests that assume stable correlation matrices.

In every one of them, the correction is the same:

1. Find the independence assumption in the governing formula.
2. Measure d from the data.
3. Replace the independent probability with the coupled probability: $p \to d \cdot p$ (or, for ensemble-style formulas, replace n with n_{eff}—the principle is the same: account for the dependence the model ignores).
4. Rerun the cost-benefit analysis with the corrected formula.

The measurement tool is the forty-line permutation test from *The Rogue Wave*. The correction is a single multiplicative factor. The implementation requires changing a standard, which requires institutional will.

This book follows *The Rogue Wave*, which proved that the error exists and measured it. This book writes the correction.

A practitioner does not need to know why d takes the value it does. The corrected formula works without a theory of the coupling, and the measurement works without one.

But there is a question the practitioner will eventually ask: given that I have measured d, and given that I have corrected the formula, is there a way to *design* the system so that d is lower? Not just to account for coupling, but to reduce it?

That question—how to build systems that are resilient to coupling, not just aware of it—goes beyond this book. Here the correction ends: measure d, put it back in the formula, and rerun the numbers.

This book gives you the correction.

There is a final thought.

Every formula in this book was written by someone who knew it was an approximation. The engineer who wrote $P = p^n$ knew that real systems are not perfectly independent. The pharmacologist who assessed drugs one at a time knew that patients take combinations. The climate scientist who divided uncertainty by $\sqrt{n}$ knew that the models share ancestry.

They used the formula anyway, because the coupling was hard to measure, and the formula without it was easy to use, and the error had not yet arrived in a form that demanded correction.

The error has now arrived. The measurement is available. The correction is one term.

The question is not whether the formula is wrong. *The Rogue Wave* settled that. The question is how long the wrong formula stays in the standard, in the regulation, in the code, in the spreadsheet,

in the textbook—after we know it is wrong, and after we know the correction.

The answer is up to the reader.

Notes on Sources

The d^n law and the permutation test (Chapters 1–2). All results are from *The Rogue Wave*, which contains the full derivation, the twelve-domain evidence, and the forty-line code. The measurement methodology is in Chapter 10 of that book; the law is in Chapter 5; the decomposition is in Chapter 3.

Supply chains (Chapter 3). The $210 billion automobile-industry revenue loss is from AlixPartners (2021). TSMC's share of more than half of global foundry revenue in 2021 is from TrendForce. The Suez Canal blockage figures (at least 369 queued ships by 28 March; about $400 million of goods delayed per hour) are from Lloyd's List. The neon concentration (two Ukrainian producers, in Odesa and Mariupol, supplying roughly half of the world's semiconductor-grade neon before the 2022 invasion; published estimates run from about 45 to 54 percent) is widely reported in the trade and general press.

Software infrastructure (Chapter 4). The 8.5 million-device figure for the CrowdStrike incident is from Microsoft's post-incident blog (July 20, 2024). The $5.4 billion direct-loss estimate for Fortune 500 companies is from Parametrix (2024). AWS market share (32%) is from Synergy Research Group (2024). The Dyn attack (October 21, 2016) and Fastly outage (June 8, 2021) are documented in post-incident reports from the respective companies.

Power grids (Chapter 5). The 52 GW capacity loss during Winter Storm Uri is from the FERC/NERC/Regional Entity Staff Report (November 2021). The 246 deaths are from the Texas Department of

State Health Services revised count (January 2022). The $195 billion cost estimate is the low end of the Perryman Group's preliminary estimate of $195–295 billion (February 2021), the figure used in *The Rogue Wave*; other estimates are far lower (NOAA's is at least $26 billion). The $5–15 billion winterization range is the estimate used in *The Rogue Wave*. Generator forced-outage rates are from NERC's Generating Availability Data System (GADS). The N-1 criterion is codified in NERC Reliability Standard TPL-001-5.

Medicine (Chapter 6). The cytochrome P450 enzyme statistics (roughly 60 to 75% of marketed drugs, CYP3A4 about 30%) are rounded from Zanger and Schwab (2013) and standard pharmacology references. The opioid–benzodiazepine co-use rates (9% in 2001 to 17% in 2013) and the 2.14 odds ratio are from Sun et al. (BMJ 2017). The adjusted hazard ratio of 3.86 for overdose death with current benzodiazepine prescription is from Park et al. (2015). FAERS data is publicly available from the FDA. Polypharmacy prevalence figures are from Kantor et al. (JAMA 2015), using NHANES data.

Construction (Chapter 7). The Champlain Towers South collapse (June 24, 2021, 98 deaths) is documented in the NIST Technical Investigation report (June 2026). The waterproofing, "extremely expensive," and "expand exponentially" language is from the Morabito Consultants report (October 2018); the $15 million remedial estimate is from the April 2021 letter to residents. Florida SB 4-D (2022) and the 2023 amendment SB 154 set the milestone-inspection schedule.

Aviation (Chapter 8). The 10^{-9} catastrophic failure rate target is from 14 CFR 25.1309 and the associated FAA Advisory Circular (AC 25.1309). ARP4761 (Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment) is published by SAE International. The β-factor model is from Fleming (1975) and subsequent updates in nuclear and aviation safety literature. The Boeing 737 MAX MCAS system and the Lion Air (October 29, 2018) and Ethiopian Airlines (March 10, 2019) accidents are documented in the Joint Authorities Technical Review (JATR,

2019) and the House Transportation Committee Report (2020).

Climate ensembles (Chapter 9). CMIP6 model count and participation are from Eyring et al. (2016). Model-dependence analyses are in Knutti et al. (2013), Sanderson et al. (2015), and Brunner et al. (2020); the correlation values in the chapter's table are illustrative. The IPCC Sixth Assessment Report (AR6) equilibrium climate sensitivity range (2.5 to 4.0°C "likely") is from Chapter 7 of the Working Group I report (2021). The Fifth Assessment (AR5) range was 1.5 to 4.5°C (2013).

Trends (Chapter 11). TSMC market share trends from TrendForce. Cloud market concentration from Synergy Research Group. Polypharmacy trends from Kantor et al. (2015). Renewable energy interconnection trends from IRENA.

General. The d^n framework, the permutation test, the decomposition, and the sign boundary are original to *The Rogue Wave*. The corrected formulas in this book—replacing p with dp in each domain's governing equation—are applications of that framework. Where a domain has its own literature on correlated or common-cause failures (the β-factor model in aviation, the effective-sample-size correction in statistics, the compound-event framework in climate science), the connection to the d^n framework is noted in the relevant chapter.

Key Formulas

Each row shows the standard formula and its corrected version. Wherever a probability p appears, the correction multiplies it by the measured coupling d; the ensemble row instead replaces the number of models n with the effective number n_{eff}.

Formula	Standard	Corrected
Pricing error	$R = 1$	$R = d^n$
Crossover	—	$n^* = \ln T / \ln d$
Effective diversification	n	n / d^{n-1}
Supply-chain survival	$1 - q^n$	$1 - (dq)^n$
Service availability	$1 - f^n$	$1 - (df)^n$
Generator outage (joint)	$\prod q_i$	$\prod (dq_i)$
Multi-drug adverse event	$1 - \prod(1 - r_i)$	$1 - \prod(1 - dr_i)$
Structural joint failure	$\binom{n}{m} p^m$	$\binom{n}{m} (dp)^m$
Redundant safety case	p^n	$(dp)^n$
Ensemble std. error	$\sigma / \sqrt{n}$	$\sigma / \sqrt{n_{\text{eff}}}$

Every corrected formula in this table reduces to the standard formula when $d = 1$. The correction is a single measured quantity: d, the per-component coupling strength from the permutation test.

Glossary

d Per-component dependence strength. Measured by the permutation test. $d = 1$ is independence; $d > 1$ is coupling.

Coupling Components are coupled when they share something—a supplier, a storm, a software bug, a power supply—that makes their failures arrive together. The opposite of independence.

Independence assumption The assumption that the chance of two things going wrong together is the product of their separate chances. Correct only when they share no cause.

d^n The pricing ratio: the factor by which an independence-based model underpredicts the real compound-event cost. Grows exponentially with the number of coupled components n.

Column shuffle A permutation that independently shuffles each column of the data matrix, destroying both the dependence between components and the order of time periods, while preserving each component's own behavior.

Common-mode failure A failure that affects multiple redundant components simultaneously, through a shared cause. Measured by d in this framework.

Coupling channel The mechanism through which components become dependent: shared geography, shared infrastructure, shared metabolic pathway, shared code heritage, shared environment.

Crossover (n^*) The number of components beyond which the model error exceeds a tolerable threshold: $n^* = \ln T / \ln d$.

Effective diversification n/d^{n-1}. The diversification benefit net of coupling cost.

n_{eff} Effective sample size. The number of truly independent members in a correlated ensemble: $n_{\text{eff}} = n/(1 + (n-1)\bar{\rho})$.

Permutation test The forty-line test from *The Rogue Wave* that measures the pricing ratio by shuffling the data to destroy one type of dependence at a time.

Pricing ratio (R) The ratio of real compound-event cost to independence-model cost. $R = d^n$.

Row shuffle A permutation that shuffles entire rows (time periods), destroying temporal persistence while preserving spatial relationships.

Lag shift A permutation that shifts each column by a random lag, destroying synchronization while preserving each component's own streakiness.

Spatial ratio (R_{spatial}) What the column shuffle destroys that the row shuffle does not: $R_{\text{total}}/R_{\text{persistence}}$.

Synchronization ratio (R_{sync}) The pricing ratio measured by the lag shift. It isolates timing alignment between components and is not the same quantity as R_{spatial}.

Phase transition (structural) In a building or other system with a collapse threshold m^*, coupling does not merely raise the probability of failure; it changes the mode from repairable (below m^*) to catastrophic (above m^*).

Total coupling The regime where $d \cdot f = 1$: all components fail when the shared dependency fails. Redundancy provides exactly

zero benefit. Occurs when a single vendor, infrastructure, or code path serves the entire system.

Index

737 MAX, 65

aviation, 65

β-factor model, 67

Champlain Towers, 57
climate models, 72
column shuffle, 16
common-mode failure, 65
coupling, 14
crossover point, 21
CrowdStrike, 35
cytochrome P450, 51

d (dependence strength), 14
d^n law, 14
diversification trap, 22
drug interactions, 50

effective diversification, 24
effective sample size, 72
ensemble, 72
Ever Given, 10

independence assumption, 14
independence ratio, 16

lag shift, 16

Loss of Load Expectation, 44

N-1 criterion, 43

permutation test, 16
persistence ratio, 16
polypharmacy, 50
power grid, 43
pricing ratio, 14
progressive collapse, 57

row shuffle, 16

software
    SLA, 35
staged rollout, 40
standard error, 73
structural engineering, 57
Suez Canal, 10
supply chain, 27
synchronization ratio, 17

total coupling, 39

Winter Storm Uri, 43