



Every Decision Added Value. The Plant Still Lost Money.

Autonomy Platform · Public Blog

P E R S P E C T I V E

© 2026 Azirella Ltd. All rights reserved worldwide.

[Read more at azirella.com/blog](https://azirella.com/blog)

TL;DR

A plant goes to a loss and every decision that put it there added value. That is how flexibility is usually spent - not by a decision but by drift, against a design envelope nobody retained, with the first honest signal arriving in the P&L hundreds of decisions late. Winter and Bullock's radical optionality is the same argument arriving in AI governance, and supply chain has had it since 1996.

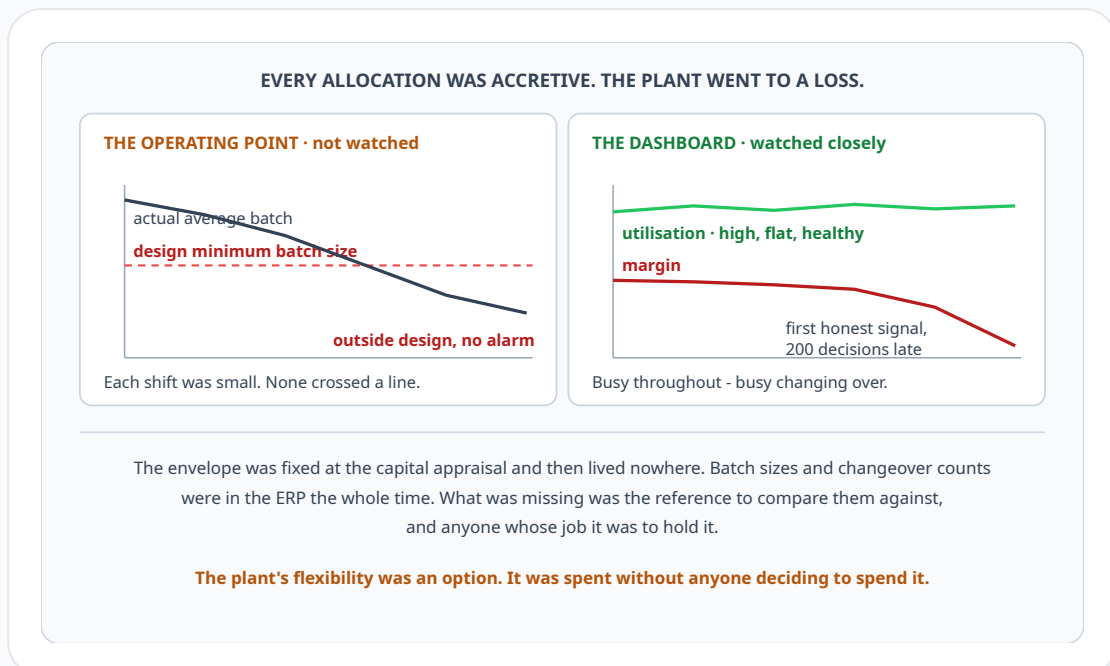
KEY POINTS

- An option is cheap only before the signal. Qualifying a second supplier takes three months normally and twelve once the shortage is public, so "buy it when we see the trigger" fails even when the warning is long. There is no later, cheaper moment.
- The method was never the gap. Risk pooling is 1979, real options on network flexibility is 1996, and Simchi-Levi's Risk Exposure Index computes consequence with no probability at all. All of it was available and the networks still ran single-source and thin.
- Count failure modes, not scenarios. The cause space is unbounded and you will never enumerate it; the effect space on your own topology is small and finite - a node gone, a link severed, lead time stretched. That works better the less imaginable the cause.

- A design is a set of assumptions about the operating point. Unless those assumptions are kept as data and compared continuously against where the network is actually running, the design quietly becomes a fiction and nobody is notified.

A large-volume pharmaceutical plant in Eastern Europe was losing money, and nobody could say why. Every product sent there had been waved through on marginal cost, because the plant had spare capacity and one more product on a paid-for asset is accretive. Two hundred of those, and the asset was doing the opposite of what it had been built to do - with no decision anywhere that anyone would defend as wrong.

I will come back to that plant. It is the clearest example I know of a cost that is invisible to every instrument pointed at it, and the route to it starts somewhere unlikely.



There is a new paper from the Institute for Law & AI, by Christoph Winter and Charlie Bullock, called *Radical Optionality: Governing Transformative AI Under Uncertainty*. Its argument is that governments faced with a technology they cannot predict should stop trying to write the right rule and start buying the capacity to write one later - information channels, evaluation capability, flexible definitions, talent depth, secure sharing. Cheap now, useful in every scenario, and impossible to assemble at short notice.

The title is doing deliberate work. It sits one word away from Kay and King's *Radical Uncertainty*, which is the book I leaned on in [Black Swans Need a Different Approach](#) to argue that most of what we call unforeseeable is merely unpriceable. The two fit together more neatly than either author probably intended, and the join is worth making explicit.

But the more interesting thing is what happens when you carry the argument into supply chain, which is where I spent my career. Because we have had this argument since the 1990s. We have the literature, the tools, the metrics and the case studies. And in 2020 we were still single-sourced, long-lead and thin, and nobody could explain why.

That gap - between having the right framework for thirty years and it not helping - is the actual subject of this piece.

What the paper proposes

Four premises. Transformative AI is possible within a decade. We have no idea what it will do. A dual-use technology with national security implications will eventually require oversight. And building the institutional capacity to provide that oversight takes years, so you cannot start when the need becomes undeniable.

From there the argument is clean, and it refuses both of the obvious positions.

Against the libertarian answer, it points out that "let the market handle it" tacitly assumes the technology will not be transformative - because nobody defends a laissez-faire regime for nuclear weapons with a straight face. Against the precautionary answer, it points out that regulation has its own expected cost, and that a rule written now will be written in ignorance.

What it rejects most sharply is the middle position, anticipatory governance - trying to predict the trajectory and regulate ahead of it. The example is California's SB 1047, which defined a covered model as one costing over a hundred million dollars to train. Sensible in mid-2024, close to meaningless within months, and amendable only by passing an entirely new bill. The paper's preference is for definitions a regulator can update, if-then commitments that fire on a trigger rather than on a date, and above all for not foreclosing: its longest argument is against the 2025 state pre-emption moratorium, on the grounds that pre-empting state law and replacing it with nothing destroys a decade of option value in a legislature that passes twenty-seven bills a year.

It is careful to distinguish itself from muddling through. Muddling preserves flexibility by inaction. This requires spending real money to preserve it - which is where the “radical” comes from.

Worth noting, since the terminology is what caught my eye: the two radicals are not the same word. In Kay and King, *radical* is a category claim - this uncertainty is of a different kind, not merely more of it. In Winter and Bullock, *radical* modifies the spending, not the optionality. Optionality as a strategy is unremarkable. Their claim is about how much it is worth paying for.

Optionality is the operational answer to the inability to price

Here is the join. Kay and King say that for a large class of consequential decisions, no meaningful probability can be attached - not because the event is unimaginable, but because there is no stationary process and no reference class to count. Winter and Bullock supply the move that follows from it: if you cannot price the event, do not buy a specific answer. Buy the ability to answer.

That is the whole of it. Optionality is what you do when the distribution is unavailable. You stop optimising against a forecast you cannot defend and start widening the set of futures you can still act in.

CAN YOU PUT A NUMBER ON THIS INSTANCE?

<p>RISK</p> <p>A stationary process. A reference class exists and can be counted.</p> <p>Yes. A distribution, and you can check it.</p> <p>Weekly replenishment.</p>	<p>RADICAL UNCERTAINTY</p> <p>You can imagine it. You may be expecting it. No probability attaches to the instance.</p> <p>No. Not now, and not with more data either.</p> <p>Pandemic. Tariff rewrite. Fire.</p>	<p>BLACK SWAN</p> <p>Could not be imagined at all.</p> <p>No, and you cannot name it.</p> <p>Rare.</p>
--	---	--

Kay is precise about the condition, and blunt about how often it holds: decisions that meet it
are a MINORITY of those to which probabilistic modelling is applied today.

And both sides reject the same two failure modes, symmetrically. Paralysis - the hard precautionary principle, which treats every change as harmful until proven safe. And false precision - which Kay and King attack as “making up numbers over and over again,

as in many Monte Carlo or similar simulations”, and which Winter and Bullock attack as a compute threshold baked into a statute. These are the same objection in two fields. A number nobody can check, defended because it is a number.

Where the paper reaches for a number it does not have

One place it slips, and it is worth naming because the slip is instructive.

To justify the spending, the paper argues that “even if there’s a 95% chance that the money spent on a given policy measure is wasted, a five percent chance of some positive impact ... would mean that the costs were justified a thousand times over in expectation.”

That is precisely the move Kay and King object to. Where did the five per cent come from? It is a guesstimate of an unpriceable thing, and multiplying it by an unbounded benefit produces a number that can justify any expenditure at all. The paper spends twenty pages explaining that you cannot price transformative AI outcomes, and then prices one.

It would be like a planner who cannot forecast a new product justifying a year of safety stock by multiplying a guessed stockout probability by the lifetime value of the customer who walks. The arithmetic is impeccable. The first number came from nowhere, and every number after it inherits that.

The honest version of the argument does not need it, which is why the slip is unnecessary. The case for a whistleblower statute or an evaluations capability is that it is cheap, it imposes almost nothing on the regulated party, and it is useful in every scenario including the boring ones. That is a dominance argument, not an expected-value one, and dominance survives the absence of a distribution. Expected value does not.

Keep that distinction. It comes back.

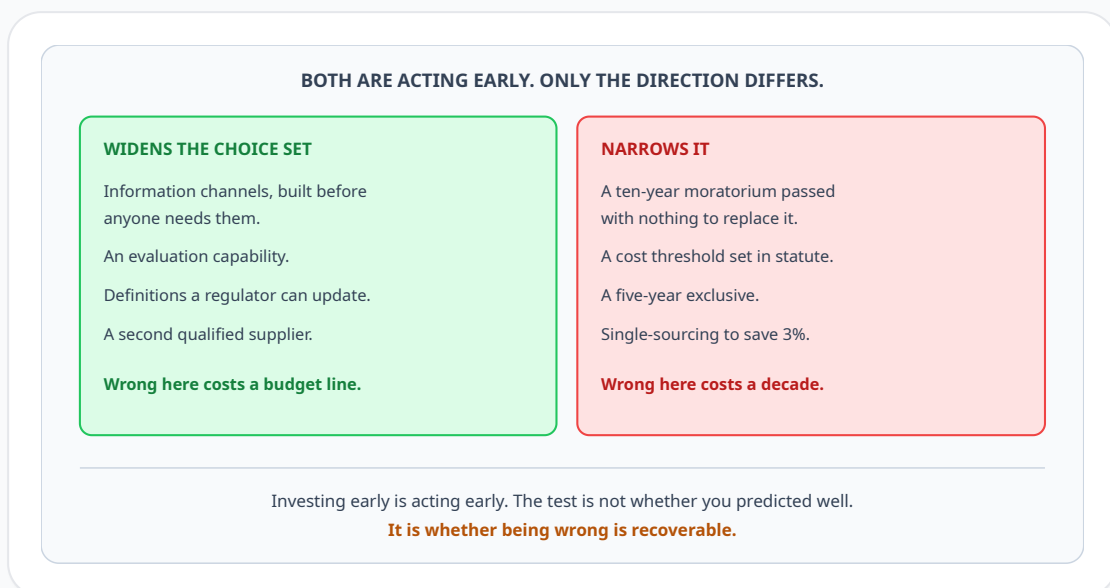
Investing early is acting early

My first instinct reading the paper was to draw a line between acting early and committing early - invest now, decide later.

That distinction does not survive contact. Building an agency is acting. Funding a reserve corps is acting. Both spend money under the same ignorance as the rule you declined to write. Calling one “preparation” and the other “regulation” is a labelling exercise.

The axis that actually carries weight is direction. Does the act widen the future choice set or narrow it? Stand up an evaluations capability and you have acted early and added options. Pass a ten-year moratorium and you have acted early and removed them. Same timing, opposite effect.

Which means the test is not prediction, it is recoverability. The paper muddies this: it attacks anticipatory governance for relying on predictions, while conceding its own programme does too - “we predict that transformative AI systems may be developed and that, if developed, they will likely be dual-use.” Everyone is predicting. The difference is that a talent reserve which turns out to be unnecessary is a wasted budget line, and a moratorium which turns out to be wrong is a decade.



Optionality has a price

The limit the paper never names: options cost money to hold.

Dual sourcing costs margin and splits volume below the price break. Reserved capacity costs cash against a call you may never make. Flexibility costs standardisation, and standardisation was earning something. Hold every option and you are as immobilised as the firm that committed - you have just spent the money on hedges rather than on the business.

All too often the response to a bad year is to buy every option at once. Dual-source everything, hold more of everything, reserve capacity everywhere, qualify a second plant for parts that have never been short. Twelve months later the cover is quietly unwound in a cost programme nobody connects to the resilience programme, and the network comes out of it thinner than it went in - having paid for resilience twice and kept none of it.

Supply chain understands this better than policy does, because we have the invoice. So the question the paper leaves open is the one that matters operationally: with no probability available, which options do you buy?

Two tests, and the one that fails

The first test is coverage: rank an option by how many distinct futures it pays in.

But stated as *scenarios*, that smuggles the problem back in. You cannot enumerate scenarios. The cause space is unbounded and the next one will not be on your list - that is the entire point of Kay and King.

The repair is to count failure modes instead. You do not need to enumerate causes to enumerate effects, because the effect space on your own topology is small and finite: a node is unavailable, a link is severed, demand steps, lead time stretches, input price moves, quality fails. That is close to all of it. A second qualified supplier covers "this node is unavailable" regardless of whether it was a fire, a tariff, a fraud, a war or something nobody has thought of. The test works *better* the less imaginable the cause, which is the exact opposite of how scenario planning behaves.

A hospital does not maintain one plan for a coach crash on the ring road and a separate plan for a fire at the chemical works. It has a major incident plan, and what triggers it is twenty people arriving at once. The causes are unbounded and nobody pretends otherwise; the thing that has to be absorbed is not.

The second test is more intuitive and it does not survive. It goes: compare build-time against warning-time. If a response takes two weeks to stand up and the signal arrives a month out, do not pre-buy it - buy it on the trigger. Pre-position only what you cannot assemble inside the warning you would actually get.

The flaw is that build-time is not a constant. It is a function of how many other people are building at the same moment. Qualifying a second supplier takes three months in a normal market and twelve once the shortage is public. Capacity you could have reserved

cheaply becomes unreservable the week everybody wants it, and allocation turns into a gaming problem - which Lee, Padmanabhan and Whang identified as one of the four causes of the bullwhip in 1997 and which 2021 demonstrated at scale.

It would be like declining house insurance on the grounds that you can always buy it when you smell smoke. The policy is cheap precisely because nothing is burning, and the moment the argument for it becomes unanswerable is the moment nobody will sell you one.

So the option reprices at exactly the moment it becomes justifiable. There is no later, cheaper moment. **An option is cheap only while nobody wants it, and that is also the only period in which you cannot build a business case for it.**

That is the mechanism behind the uncomfortable symmetry: uncertainty is highest when optionality is cheapest, and both decay together as the event approaches.

What survives is a ranking with no time axis and no probability in it: consequence-if-absent, which is a property of the topology and computable from the graph you already have, weighed against cost-to-hold, which you know exactly. The price of the cover while nobody else wants it is the honest valuation. There is no better one coming.

We have known all of this since 1996

Here is where a supply chain reader should be getting impatient, and should.

Everything above is a restatement of a literature that is thirty years old or older, and in places the original is sharper than my version of it.

- **Failure modes rather than scenarios** is FMEA, formalised in US military procurement in 1949 and a standard discipline in automotive quality by the 1990s. Severity, occurrence, detection - and the severity column does not need the occurrence column to be useful.
- **Consequence without probability** is David Simchi-Levi's Risk Exposure Index, developed with Ford and published in *Harvard Business Review* in January 2014. Compute each node's Time-to-Recover, compute the financial impact of operating without that node for that long, and rank on the result. The method deliberately does not estimate likelihood, on the explicit grounds that you cannot estimate the probability of a rare disruption. That is the rule I spent this essay arriving at, published twelve years ago, and it won the Wagner Prize the same year.

- **Real options on network flexibility** is Huchzermeier and Cohen, *Valuing Operational Flexibility Under Exchange Rate Risk*, Operations Research, 1996.
- **Risk pooling** is Eppen, 1979. **Postponement** is older than that. **The resilience vulnerability map** is Sheffi and Rice, 2005.

So the method was never the gap. Which makes the real question not “what should we have done” but “why did having all of this since 1996 not help in 2020”.

Why it did not help

Three answers, and not one of them is analytic.

It was done once. A network design study is a project. It runs for three months, produces a recommendation and a deck, and then the topology keeps moving while the analysis does not. Two years later the answer is describing a network that no longer exists, and nothing announces this. A network study has the shelf life of a photograph and we keep filing it as though it were a map.

It was disconnected from the operating plan. The resilience study and the weekly planning run were different systems, different data and different owners. So the mitigation was never costed against the margin it displaced in the terms the business actually argues in, and it lost every budget round to a number that was in the plan.

The load-bearing input was surveyed, not measured. Time-to-Recover - the number the whole calculation rests on - is typically collected by asking suppliers how long they would take to recover. Asking a supplier how long they would need to recover from a fire is like asking a builder how long the job will take. The answer is not dishonest, it is aspirational, and nobody ever goes back and scores it against what happened. Self-reported, unvalidated by anything, and stale the moment it is collected. If that number is carrying the analysis, it is the one thing that should be derived from observed behaviour rather than from a questionnaire.

Three different failures with one shape: **a value captured once at design time and never re-derived while the world moved underneath it.** Which is, exactly, SB 1047's hundred million dollars.

Back to the plant

So, back to the plant. It came to me from someone who had run operations for a generic pharmaceutical manufacturer. Large volume, built for long campaigns of high-runner products - big batches, few changeovers, a heavy fixed base amortised across volume. Over the years it became the place the tail of the portfolio went: low-volume products, slow movers, the long thin end of the range.

Nothing about the loss is mysterious once you see the shape. What is worth dwelling on is how thoroughly it hid, because every part of that is general.

No single decision was wrong. Each tail product was allocated there because the plant had spare capacity, and each was waved through on marginal cost - one more product on an asset that is already paid for looks accretive, and on a marginal basis it genuinely is. Two hundred of those and the plant is running small batches with constant changeovers on equipment designed for the opposite, carrying a full fixed base against work that was all costed at the margin.

Nothing breached a threshold. This was drift, not an event. There is no day on which the plant left its design envelope, so there is no alert that could have fired on a limit.

The monitored metric looked healthy. Utilisation stayed high throughout. The plant was busy. It was busy changing over. High utilisation with collapsing effective throughput is the signature of exactly this failure, and if utilisation is what you watch, you see a plant in good health right up to the end. It is the kitchen that is flat out all service because it is cooking forty different dishes for parties of two: every hob lit, nobody idle, and barely a cover served.

The decision and the consequence were in different places. The allocations were made in portfolio and supply planning. The cost landed on a plant P&L whose owner never saw the decisions and could not have contested them.

The first honest signal was financial. By the time the margin moved, the cause was two hundred decisions back and attributable to none of them - which is why nobody could explain it. There was nothing to find at the level they were looking.

A design is a set of assumptions nobody keeps

That plant had an operating envelope: minimum economic batch, campaign length, changeovers per week, how many distinct products the line was meant to carry. Those numbers existed. They were in the capital appraisal, and in the head of whoever sized

the asset. And then they lived nowhere.

Nobody was missing data. Batch sizes and changeover counts were in the ERP from the first day. What was missing was the reference to compare them against, and an owner for the comparison.

So, the general form:

A design is a set of assumptions about the operating point. Unless those assumptions are retained as data and compared continuously against where the network is actually running, the design quietly becomes a fiction - and nothing notifies you.

It would be like a building with a stated floor loading that nobody ever checks against what is actually stored on it. Nothing collapses on the day the limit is passed. The structure simply stops being the one that was signed off, and carries on looking exactly as it did.

Note what this is *not*. It is not an argument for more metrics. Plants do monitor changeover time and batch size. They monitor them as performance against an improvement target - “are we doing this quickly” - which is a different question from “should we be doing this here at all”. The measurement existed. The comparison did not.

And now connect it back, because this is the same failure three times: the network study on a snapshot that rotted, the Time-to-Recover surveyed once and never re-derived, the statutory threshold written in 2024 for a world that changed in January 2025. A design parameter fixed at a moment and then left to be overtaken.

Which means Winter and Bullock are right about the remedy - regulatory definitions over statutory ones, precisely because they can be revisited - and the supply chain evidence says the remedy is not sufficient. A definition that *can* be updated and is not is indistinguishable from one that cannot. Something has to be watching the gap between the design and the operating point, continuously, and telling somebody.

Where the argument is uncomfortable

Two places.

Some of it was a bet, not a failure. For a good number of firms, thin and single-sourced was a position that paid for twenty years and lost once. Calling that a monitoring failure is hindsight dressed as analysis. The drift reading holds where the

operating point moved without anyone choosing it - which is the common case, and is what the pharma story is - but not where somebody looked at the exposure, understood it, and took the margin anyway. Those are different errors and only one of them is an error.

Monitoring does not win the argument on its own. The binding constraint in 2020 was never analytic capability. It was that nobody will pay for cover while nothing is wrong, and every incentive in the business is pointed the other way. Making the drift visible does not change that - it changes *when* the conversation can happen. And given that the option is only affordable while nobody wants it, the window in which that conversation is both possible and useful is precisely the window in which the business case is weakest. The monitoring buys you the chance to have an argument you will probably still lose. That is a real improvement over not knowing, and it is less than it sounds.

What this asks of you

Four things, in increasing order of difficulty.

Keep the envelope. The design assumptions - minimum batch, campaign length, lane transit assumptions, the demand mix the network was sized for, the supplier count each node was meant to have - are data. Write them down somewhere a machine can read them. Right now, for most networks, they are in a five-year-old appraisal and a retired engineer's memory.

Compare continuously, and report the drift rather than the breach. The failure is gradual, so a threshold alarm fires too late by construction. What you want reported is the distance between the design point and the operating point, moving, with a trend - not a limit that was crossed.

Measure Time-to-Recover rather than asking for it. The number that carries the entire exposure calculation should come from observed behaviour - actual recovery from actual disruptions, actual qualification lead times from your own history - not from a supplier questionnaire returned by somebody incentivised to answer well.

Rank on consequence against cost-to-hold, and stop waiting for a probability. It is not coming. Criticality is computable from the graph you already have, with no event prediction in it. The cost of the cover you know to the pound. That is a complete decision rule, and it has been one since 2014.

The paper's deepest assumption is that institutional capacity produces better decisions. Supply chain is the natural experiment for that claim, and the result is equivocal: we had the capacity, the literature and the tools, and the networks still ran thin. Not because the analysis was wrong, but because the analysis was a project rather than a process, and the design it was defending had quietly stopped describing the network.

Optionality is almost never spent by a decision. It is spent by drift, one individually accretive allocation at a time, while the dashboard shows a plant that is pleasingly busy.

Sources. Christoph Winter and Charlie Bullock, *Radical Optionality: Governing Transformative AI Under Uncertainty*, Institute for Law & AI, 23 April 2026. John Kay and Mervyn King, *Radical Uncertainty: Decision-Making Beyond the Numbers*, W. W. Norton, 2020. David Simchi-Levi, William Schmidt and Yehua Wei, *From Superstorms to Factory Fires: Managing Unpredictable Supply-Chain Disruptions*, Harvard Business Review, January-February 2014, for Time-to-Recover and the Risk Exposure Index; the underlying Ford work took the 2014 INFORMS Daniel H. Wagner Prize. Arnd Huchzermeier and Morris Cohen, *Valuing Operational Flexibility Under Exchange Rate Risk*, Operations Research 44(1), 1996. Gary Eppen, *Effects of Centralization on Expected Costs in a Multi-Location Newsboy Problem*, Management Science 25(5), 1979, 498-501. Yossi Sheffi and James Rice, *A Supply Chain View of the Resilient Enterprise*, MIT Sloan Management Review 47(1), Fall 2005. Hau Lee, V. Padmanabhan and Seungjin Whang, *The Bullwhip Effect in Supply Chains*, Sloan Management Review, Spring 1997, 93-102, where shortage gaming is one of the four named causes. FMEA originates in US military procedure MIL-P-1629, *Procedures for Performing a Failure Mode, Effects and Criticality Analysis*, 9 November 1949.

See Autonomy in action

Walk through how Autonomy models, executes, monitors, and governs supply chain decisions with autonomous AI agents.

[See It Live](#)