



Black Swans Need a Different Approach

Autonomy Platform · Public Blog

PERSPECTIVE

© 2026 Azirella Ltd. All rights reserved worldwide.

Read more at azirella.com/blog

TL;DR

Almost nothing we call a black swan is one. Kay and King's radical uncertainty names the real category: events you can imagine, may even be expecting, and still cannot attach a probability to. That reclassification takes away the comfort, because if a pandemic was anticipated in kind the question stops being why did we not predict it and becomes why was the network not built to survive it.

KEY POINTS

- Being unable to price the event says nothing about your ability to prepare the response. A twin does not tell you how likely a scenario is, it tells you how a response performs given one, and only the second question has a reference class.
- Criticality is a property of the topology, computable today from the graph you already have with no event prediction in it. It was legible the whole time and nobody was reading it.
- The error nobody names costs more. Treating a mystery as a puzzle produces spurious precision; treating a puzzle as a mystery convenes eleven people for ninety minutes on a number a machine should have produced overnight, and it does not look like a mistake, it looks like diligence.
- Distribution-free is not assumption-free. Conformal prediction assumes exchangeability, which is Kay's stationarity condition restated as a testable property of a dataset. We found eight versions of that overclaim on our own site and rewrote all of them.

Every supply chain shock of the last fifteen years has been followed by the same conclusion: we need better forecasts. The tsunami, the volcano, the canal, the pandemic, the tariffs. Each time, the lesson taken away was that we should have seen it coming and next time we will.

We will not. But the reason is not the one usually given, and what follows from it is not despair.

Start with the word. A black swan is an event that could not have been *imagined* before it happened - the settlers who named it had no concept of a swan that was not white. By that standard, **almost nothing on the list above is a black swan**. John Kay is blunt about the biggest one:

*"A global pandemic of this kind was not, although many people have described it in this way, a 'black swan' - something which could not have been imagined before it happened... We wrote in our book that 'we must expect to be hit by an epidemic of an infectious disease resulting from a virus which does not yet exist'. **A pandemic was a widely anticipated event**, but it makes no sense to ask, either before or after that event, 'what was the probability that one will break out in Wuhan in December 2019?'"*

Read that twice, because it contains two claims and only the first is comfortable. The first is that you were not ambushed. The second is that being right about the *kind* of event bought you no ability whatever to put a number on the *instance*.

Risk, uncertainty, and the distinction that got buried

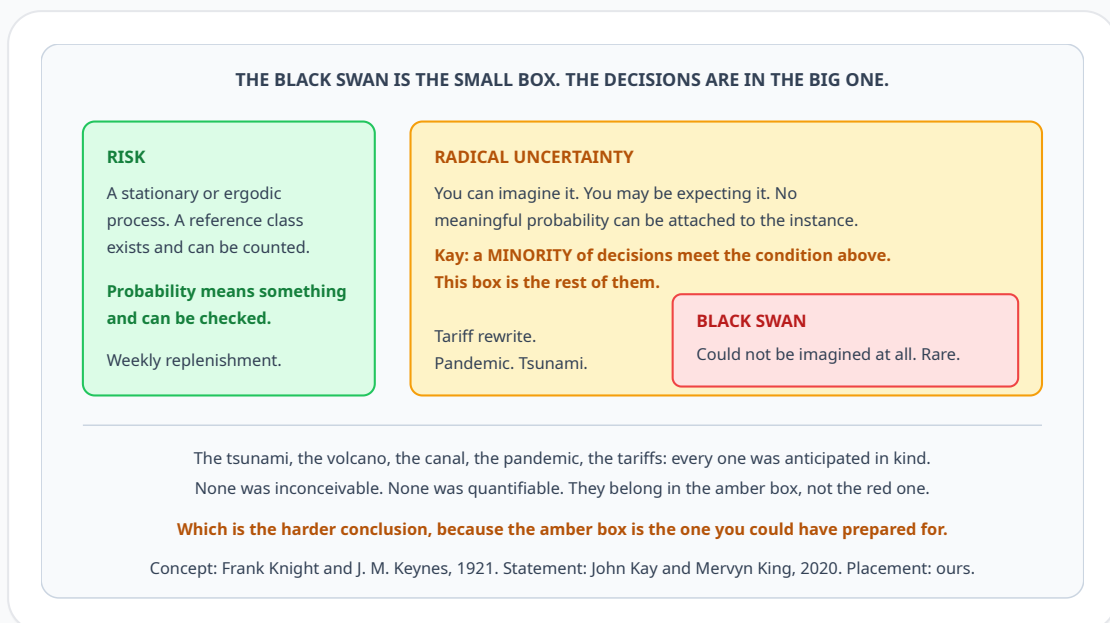
In 1921 two books drew the same line: Frank Knight's *Risk, Uncertainty and Profit* and Keynes's *A Treatise on Probability*. Keynes put it better, in 1937:

*"By 'uncertain' knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty... The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention... **About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know.**"*

Kay and King's *Radical Uncertainty* picks that line back up. Some uncertainty is **resolvable** - go and find out. Some can be **usefully characterised probabilistically**, and Kay is precise about the condition: "where outcomes are known to be the product of some stationary or ergodic process." Everything else is **radical uncertainty**. Then the sentence that should be pinned above every planning system in the industry:

*"these situations, in which probability distributions can be deduced from knowledge of the underlying determinants of outcomes or by observation of a long or extensive historic series of these outcomes, **are a minority of those to which probabilistic modelling is today applied.**"*

Not a rarity. A minority. And they do not reject Taleb so much as **contain** him: a black swan is where you cannot imagine the event at all, an extreme and special case. Radical uncertainty is the much larger territory where you can imagine the event perfectly well, may be actively expecting it, and still cannot say how likely it is or when.



Calling something a black swan has become a way of ending a conversation. It says the event was outside anybody's reach, so nobody could have done anything, so there is nothing to learn. The reclassification takes that comfort away. If the pandemic was anticipated in kind and merely unquantifiable in timing, the question is not "why did we not predict it" but "why was the network not built to survive a thing we knew would eventually happen."

That is a harder question and a much more useful one, and the rest of this piece is about answering it rather than about the pleasures of epistemic humility.

You cannot price the event. You can prepare the response.

This is the part that gets lost, and losing it turns a good argument into a counsel of despair. **Being unable to attach a probability to the event says nothing about your ability to prepare for it.** Those are different objects, and only one of them was ever unavailable.

The line is narrow and worth drawing exactly. **A digital twin does not tell you how likely a scenario is. It tells you how a response performs given one.** Only the second question has a reference class.

You cannot calibrate “will there be a tsunami.” That instance happens once, and Kay is right that no amount of re-running produces a probability for it. But “when a supply source vanishes, does lateral rebalancing recover fill rate, and by how much” is a question that recurs across every disruption, every drill and every scenario you care to generate. It is a small world nested inside a large one: **the antecedent is radically uncertain, the conditional is not.**

Three things follow, and none of them requires forecasting anything.

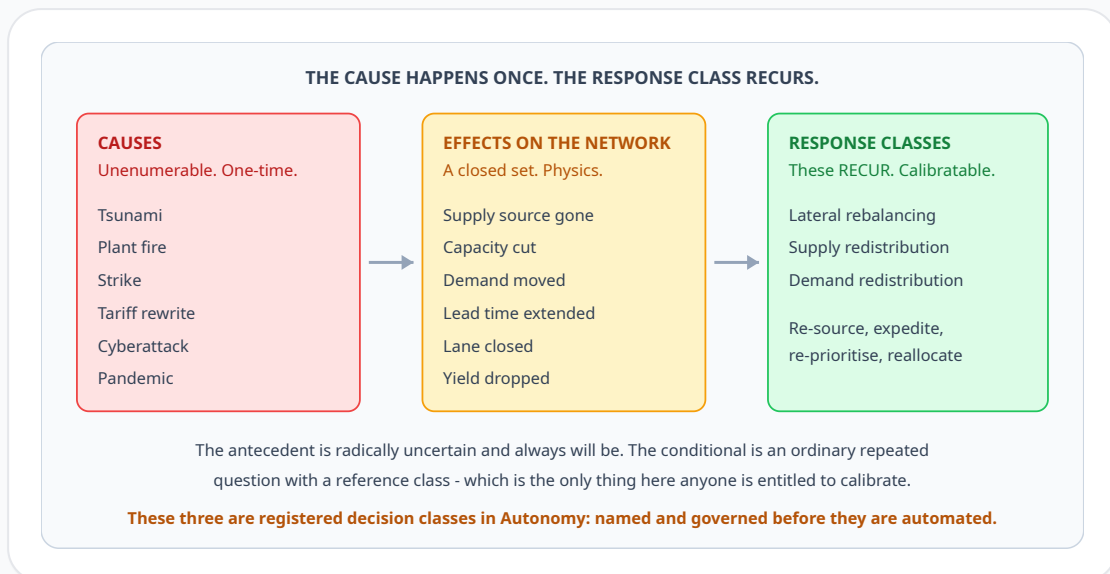
You know which nodes would hurt, before anything happens. The structure of the network is present tense, not future tense. Which components are single-sourced, which suppliers you are concentrated in, which sites sit on a path with no alternative, which node’s failure disconnects the most demand - all of that is computable today, from the graph you already have, with no event prediction anywhere in it. Criticality is a property of the topology. It does not become knowable only after the shock arrives; it was legible the whole time and nobody was reading it.

You can pre-validate the responses, not just the plan. For a class of disruption - a supplier lost, a lane closed, a plant down - the recovery options can be evaluated and ranked in advance and kept ready. When one arrives, the question becomes which prepared answer fits rather than how fast can we re-solve from nothing. That is not prediction. It is the difference between a fire drill and a fire.

And you can run it on a Tuesday. If the machine can evaluate alternative configurations against a scenario in minutes, you do not have to wait for the tsunami to ask the question. What happens if this supplier vanishes, what would we do, what would it cost, and is there something cheap we could change now that would make the answer better. That is a **drill**, and it is the same computation as the response - which means the response has been rehearsed rather than improvised.

What none of this does is rank the scenarios by likelihood, and the tempting version is exactly that: hand each scenario a probability so the planner can rank them and feel the ranking is objective. That is precisely the move Kay and King spend a book demolishing, and they are right. A probability on a one-time event is unfalsifiable, and an unfalsifiable number placed next to a real one borrows its authority.

What you can legitimately put beside a response is **how it holds up across the scenarios**, which is a statement about the responses and not about the world. Different object, and it has a forty-year pedigree.



Small worlds, and the error that runs both ways

The distinction was buried deliberately - Ramsey, de Finetti, Savage, then Friedman, who wrote that Knight's distinction is one "I do not believe... is valid". The stakes explain the effort: without universal subjective probability, as Kay notes, "modern portfolio theory, the capital asset pricing model, and much of rational expectations based macroeconomics would fall away." The technical objection - refuse to hold consistent probabilities and you can be made to accept a losing set of bets - gets the answer that carries the whole book:

*"that is precisely what Keynes and Knight - and us, following them - do deny. **Rational people do not bet on most things. The rational response to uncertainty is generally to say 'I don't know'.**"*

Savage himself limited his axioms to **small worlds**: settings where every outcome can be enumerated and nothing outside the list can occur. The axioms were then applied everywhere. Kay and King's own pairing is **puzzles** and **mysteries**: a puzzle has a

definite answer, known to be findable, and you know when you have it; a mystery is ill-defined, and you may not know the answer even after the event.

The error everyone recognises is **treating a mystery as a puzzle**. A re-shoring decision run through a model that produces a net present value to four significant figures, when the inputs are guesses about a political process nobody can enumerate. Kay and King's term is **spurious precision**: "When no meaningful quantification is possible, algebra can provide only spurious precision, while at the same time the language becomes casual and sloppy." That second clause is the one people miss. The precision does not merely fail to help. It *substitutes* for the harder thinking.

The error nobody names costs more. **Treating a puzzle as a mystery** is what an organisation does when it convenes eleven people for ninety minutes to decide a replenishment quantity that has a defensible answer, computable from data the company already owns, and that will be re-decided next week regardless. The cost is not the ninety minutes. It is that those eleven people were the only resource in the building capable of handling the re-shoring question, and they spent the morning on a number a machine should have produced before they woke up. **This failure does not look like a mistake. It looks like diligence.**

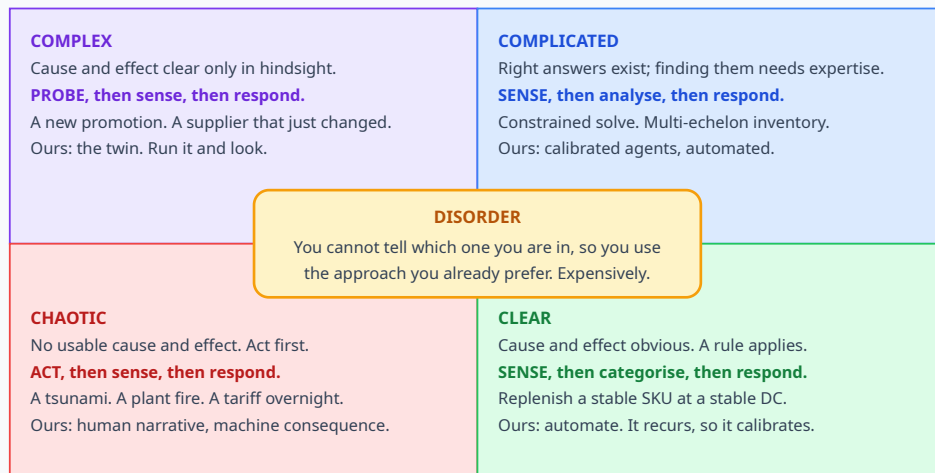
Both errors have one root: nobody asked which kind of thing was on the table.

Cynefin: which move comes first

Kay and King tell you what kind of uncertainty you face. They are deliberately unhelpful about what to do next - they refused to end the book with a numbered list of lessons, on the grounds that the genre sells well and misleads. Fair as philosophy. Less useful on a Monday.

Dave Snowden's Cynefin framework does the work the philosophy does not. Built at IBM in 1999, it sorts problems not by difficulty but by **what kind of response works**. The two frameworks are not rivals, and the way they fail to line up is the most useful thing about running both: Cynefin sorts by what kind of response works, Kay and King by whether the process underneath is stationary. The gap between them is exactly where the expensive mistakes live.

THE SAME FOUR DOMAINS, WITH A SUPPLY CHAIN IN THEM



The response verbs are the point. A domain does not tell you the answer; it tells you what kind of move comes first, and getting that wrong is not a small error. Framework: Dave Snowden, Cynefin, 1999. Mapping: ours.

Read the verbs rather than the labels. In **Complex**, cause and effect are visible only in hindsight, so analysis has nothing to bite on. You **probe** first, with a small safe-to-fail experiment, then sense, then respond. In **Chaotic**, there is no usable cause and effect at all and the pressing need is to establish some. You **act** first, then sense, then respond.

And “act” is far more specific than it sounds, because *act before you understand* reads as a licence to flail and is nothing of the kind. **The act is to impose an assumption** - to state, deliberately and in a form a model can be run against, something you do not know to be true. That is not a forecast. It is a working premise, chosen because it is stateable and because the network can be evaluated against it in minutes.

Kay and King have a name for that object.

“What is going on here?” - the reference narrative

Having spent half a book dismantling the probabilistic apparatus, they have to say what replaces it. Their answer is a question:

*“the rational individual will stand back and ask ‘what is going on here?’... **Abductive reasoning, or inference to best explanation, is how we make sense of an imperfectly understood present and imperfectly known future. Frame a reference narrative, a realistic scenario - business strategy, policy trajectory, personal financial plan. Risk is then an imaginable event which threatens to derail the achievement of that scenario.**”*

That last sentence is a definition of risk a planner can use. Not variance. Not a standard deviation of demand. **Risk is what derails the story you are operating to.**

And narrative is not a consolation prize for people who cannot do the maths:

*“They think in terms of narratives rather than numerical probabilities, **not because they are irrational**, but because a narrative approach has proved the best way of managing the ill-specified mysteries that have confronted them over the millennia...”*

The planner who distrusts a confident model output and reaches for a story is not being soft. They are applying the right instrument to a large world, and a system that treats that reflex as an error to be trained out is removing the one thing that works there.

Repetition is how you tell which world you are in

Kay's condition for legitimate probability is “a stationary or ergodic process.” True, and useless at a desk. Nobody in an operations meeting is going to test for ergodicity.

There is a proxy, and it is unglamorous and reliable: **does this decision recur?**

A decision that recurs builds a reference class - many instances of the same choice, with outcomes you can count. That is what lets you attach a probability and later check whether it was any good. You said ninety percent; did it happen ninety percent of the time? A decision that happens once has no reference class and never will, so any probability attached to it is unfalsifiable, and unfalsifiable confidence is the most dangerous thing you can put in front of a decision-maker.

So repetition is the **diagnostic**:

- **Recurrs often, resolves fast.** Small world. A reference class exists, coverage is measurable, a calibrated Achievability means something. Automate it.
- **Recurrs, resolves slowly.** Still a small world in principle, but the evidence arrives at the pace of the horizon. Thin ice. Inform rather than automate.
- **Happens once.** Large world. There is no reference class, there never will be, and the honest response is a person, a narrative, and a machine that computes consequences fast.

Two traditions arrive at the same line. Kay and King get there from philosophy. Decision-systems practitioners got there from engineering: at DecisionCAMP this year, Jacob Feldman opened by defining the field's scope as operational business decisions - **not**

strategic, **not** one-time, day-by-day and repetitive. You cannot build a decision service for something that happens once.

The small world, done honestly

Most of a supply chain lives here, and this is where we automate. But “it recurs” is a claim about the world that can stop being true, so the interesting engineering is not the automation. It is knowing when you have left.

The mechanism is a calibrated band rather than a point estimate. Conformal prediction gives a finite-sample coverage guarantee without assuming your data are Gaussian or that the model is correctly specified. Which brings me to a sentence that was, until recently, published on our own site:

“When the system says ‘90% confident,’ it means the true outcome falls within bounds at least 90% of the time, guaranteed, regardless of the underlying distribution.”

Every clause is true. The sentence is still doing something I am not comfortable with.

Distribution-free is not assumption-free. The assumption it does make is *exchangeability*: that the data it was calibrated on and the thing being predicted come from the same process. “Regardless of the underlying distribution” invites you to hear “regardless of anything,” and omits the condition carrying all the weight. We found eight versions of that claim across our own pages. All of it is rewritten.

I flag it because of what exchangeability *is*. It is Kay’s stationarity condition, restated as a property of a specific dataset - the small-world boundary written down in a form you can test. A system that quietly assumes exchangeability is asserting it is in a small world without checking.

Coverage, though, is measurable after the fact. You promised ninety percent; you count how often the realised value fell inside. While the process holds, empirical coverage tracks the promise. When it stops, the two separate visibly, in a number, on a schedule.

The band cannot survive a regime change. It can report one - which is the difference between a system that leaves the small world quietly and one that tells you it has. Contrast a model’s own confidence output, the number most AI products surface: checked against nothing, high nearly all the time, and highest in exactly the novel situations where it should not be.

So the check runs on a cadence, keyed to the horizon of each band rather than the planning bucket - a band for a step twelve weeks out cannot be scored until twelve weeks have passed.

That carries a concession. At the strategic end, a two-year band cannot be calibrated from one network's own history, because the history does not exist and never will. It has to come from pooling across comparable products and regions, or from the twin - and the band should tell you which, because those are different assumptions and you are entitled to know which one you are relying on. The further out you go, the more the "small world" claim is something we are asserting rather than something we have measured.

And I should be exact about tense, because this piece has spent several thousand words on people who are not. The loop that turns sustained drift into a retrained policy is **built and wired; it is not yet fed**. Every link depends on decisions whose outcomes have actually resolved, and a decision made today resolves on the horizon it was made for. The honest description is a mechanism waiting on evidence, not a loop currently closing. Anyone who tells you their learning loop is closed on a network they deployed this year is describing a design, and should be asked how many decisions have resolved.

There is a radical-uncertainty reason for the caution. **A regime break is the moment of minimum evidence about the new regime**. Retrain the instant you detect drift and you are fitting noise with great conviction. The honest immediate response is the band widening on its own - which is the system saying it is less sure than it was - and the retrain follows once there is enough of the new world to learn from.

Report, widen, then relearn. In that order, on a clock, whether or not anybody was watching.

There is a related discipline about simulation, and it is the sentence in *Radical Uncertainty* that lands hardest on anyone selling this kind of software:

*"making up numbers does not help the understanding or management of radical uncertainty. **Still less does it help to make up numbers over and over again, as in many 'Monte Carlo' or similar simulations. Different guesstimates of the same thing do not constitute a probability distribution.**"*

We reached the same conclusion from the engineering side. The architectural rule in our own codebase is that there is no Monte Carlo in planning: uncertainty is carried by the conformal band, which is checkable against outcomes, not by sampling a model repeatedly, which is not. Sampling your own assumptions a thousand times produces a tight distribution over your assumptions, and its tightness is actively misleading.

Chaotic: a reference narrative, and a machine to compute it

In 2011, when I worked at Kinaxis, I watched a customer respond to the Tōhoku tsunami. What they did contains no probability at all, and it is both Cynefin's chaotic response and Kay and King's reference narrative, executed almost to the letter.

Nobody forecast the tsunami. Nobody assigned it odds. What the planner did was **act**: they built a scenario on one assumption - *any supply from Japan not already confirmed on a boat will never arrive*. That sentence is a reference narrative in exactly King and Kay's sense. It is not a prediction and nobody claimed it was likely. It is a realistic scenario, stated so that everything else can be worked out from it, and it converts a fog into a question with an answer.

Then they **sensed**: computed the business impact across the network. Then they **responded**: reallocated available finished goods and raw material, and found alternative sources for the components feeding the demand the reallocation had hurt most.

The machine's contribution was not prediction. It was computing the consequence of a human-supplied narrative across a whole network fast enough to act on. That is the correct division of labour.

They could and did build several response scenarios and compare them side by side, fast. I am describing 2011-era tooling and make no claim about what that product does now. The gap then was that nothing came back with a **robustness** measure: no way to ask which response still works across the range of ways this could go, rather than under the single line you typed. So several fast, coherent, mutually exclusive answers come back, and nothing ranks them. The ranking falls to whoever is in the room - which is not a criticism of the planner, but of giving them nothing to put beside their judgement.

A black swan is unintended supply chain design

When the tsunami took out a tier-two supplier, it **redesigned that customer's supply chain**. Not figuratively. The set of nodes changed, the set of lanes changed, the sourcing structure changed. The network that existed on the twelfth of March was not the network that existed on the tenth. Nobody chose it; it happened anyway.

So the chaotic response is not a new discipline. **It is supply chain design, executed under time pressure, on a network you did not choose.** And supply chain design is a settled field with a settled method: propose alternative configurations, evaluate each against a set of scenarios, and choose on **robustness across the set** rather than optimality against any one of them. You do not forecast which scenario will happen - that is the entire point - you find the configuration that is least bad across the ones you can imagine.

Which is, almost word for word, Kay and King's prescription:

*"a sound strategy, policy or plan is one which is **robust and resilient** to the risks that a radically uncertain future will present. **This strategy, policy or plan will probably not be optimal - radical uncertainty without probabilities means you do not know what is optimal, and probably won't know after the event what would have been optimal.** As Herbert Simon emphasised long ago, real decision makers do not optimise but **satisfice**."*

Design engineers have never needed telling. Note what design has always had to accept: **there is no history for a network that does not exist yet.** You cannot calibrate a proposed DC from its own past. Every serious design tool ever built therefore runs on simulation, and nobody has called that false precision, because in design there is no alternative instrument. That is a weaker claim than it sounds and worth keeping weak: it makes simulation legitimate where history is absent. It does not make it accurate, and it earns no licence back on the planning side, where history exists and can contradict it.

THE SAME EXERCISE, WITH THE CLOCK SPED UP		
	DELIBERATE DESIGN	BLACK SWAN
Who changed the network	You	The world
How long you have	Months	Hours
Starting point	Today's network	Today's network, minus one
What you evaluate	Alternative configurations	Alternative responses
Against what	A scenario set	One scenario, already true
How you choose	Robust across scenarios	Best against the one you have
Every concept transfers. What does not transfer is the leisure.		

This is why the twin sits where it does. Planning is repetitive and has history, so you calibrate from history - and letting a simulator near the plan of record there is exactly the invented precision this argument objects to. Design has no history by construction, so simulation is the only instrument there has ever been. A black swan is design arriving unannounced, which puts it on the design side of that line.

Design is what you do before. Response is what you do after. **If they were different machines you would only ever have the second one.**

Where the argument is uncomfortable

Two places, and I would rather name them than have them found.

The first is theirs. Their prescription - modularity and redundancy, robustness over efficiency - runs into a question the book does not answer. Redundancy costs money now and pays off on a date nobody can name. Pressed on this, King offered regulation, and the observation that a chief executive holds the job for five years and gets all the cost of resilience and none of the benefit. Both true; neither is a market mechanism, and neither helps the operations director whose competitor is running leaner this quarter. Anyone telling you resilience pays for itself should be asked over what horizon and out of whose budget.

The second is ours, and it is what the argument above obliges us to publish rather than assert. If robustness is bought by the width of the training distribution, then the way to know is to withhold a disruption type from training entirely, score the policy against it one type at a time, and publish the degradation curve. We have built the mechanism to run that. We have not yet run the full sweep. Until we have, “robust” is a design argument and not a measurement, and I would rather say so than let the word do work it has not earned.

Which is the shape of question worth asking of anyone, ours included. Not “is it robust”, because everyone answers yes. **“What would you have measured to know, and what did it say?”**

What this actually asks of you

Not a rebuttal of anyone. A habit, and it is cheap.

Before the model, before the meeting, before the escalation, ask Kay and King’s question: **what is going on here?** And then the operational form: does this decision recur, and does it resolve fast enough to check?

If yes, you are in a small world. A machine should be doing it, and you should be measuring whether the machine was right - which also means watching for the day the small world stops being one, because it will.

If no, no probability will help you, and anyone offering one is selling you a number that cannot be checked. What you need instead is a reference narrative, a fast way to compute its consequences, and a person whose judgement you trust. Then you choose on robustness across the stories you can imagine rather than optimality against the one you happen to believe.

And you do the computing **before** the event, not during it, because the machine that answers “what would we do” on a Tuesday afternoon is the same machine that answers it on the morning everything breaks.

The failure mode is not being wrong about the answer. It is never having asked the question, and reaching for the response you always reach for.

So stop saying “black swan”. It is almost always false and it is usually an excuse. The pandemic was anticipated. The tsunami was anticipated. What was never available was the number - and the number was never the thing you needed.

Sources. John Kay and Mervyn King, *Radical Uncertainty: Decision-Making Beyond the Numbers*, W. W. Norton, 2020; John Kay, *Radical Uncertainty*, SUERF Policy Brief No. 22, August 2020, which is the source of the Knight, Keynes, Friedman, Monte Carlo, reference-narrative and robustness quotations; Kay and King in conversation with Michael Mainelli, Z/Yen Financial Services Club, 15 April 2020, for the black-swan-as-special-case and resilience-incentive material; Kay and King on *EconTalk*, 2020, where the narrative passage quoted here was read aloud from the book; Kay and King, *The radical uncertainties of coronavirus*, Prospect, 2020, for “spurious precision”. Frank Knight, *Risk, Uncertainty and Profit*, 1921. J. M. Keynes, *The General Theory of Employment*, QJE, 1937. Nassim Nicholas Taleb, *The Black Swan*, 2007. Dave Snowden, the Cynefin framework, 1999; Snowden and Mary Boone, *A Leader’s Framework for Decision Making*, Harvard Business Review, 2007. Jacob Feldman’s scope-setting opening at DecisionCAMP 2026. OpenAI et al., *Learning Dexterous In-Hand Manipulation*, 2018, for domain randomisation. Vladimir Vovk, Alexander Gammerman and Glenn Shafer, *Algorithmic Learning in a Random World*, Springer, 2005, for the conformal coverage guarantee and its exchangeability condition.

See Autonomy in action

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

[See It Live](#)