



● PAPER

The Resolution Gradient

Michael Kontorovich · 1 October 2026 · 15 min read

IN BRIEF

Why might learning drift into the machinery when no one intends it? Machine capability can be bought quickly, human capability has to be grown, and so each sensible local decision tilts the same way. The paper names this the resolution gradient and describes the ratchet that locks it in. It takes the question left open in *Resolution, Not Revolution*, where learning should live, and answers with a return path: what a system learns has to come back and change the people in it. It then asks who will pay to keep that path open.

Resolution, Not Revolution left open a disturbing possibility.

As human interdependence increases, problems can arise that cannot be adequately resolved at the level that produced them. One response is higher-order organisation: the We acquires greater capacity to perceive, remember, coordinate, learn and act as a whole.

But that does not tell us where this new capacity will reside.

We distinguished two possibilities.

A_{human} — collective autonomy resides in the human network.

$A_{\text{connection}}$ — collective autonomy resides in the connective machinery.

Both can resolve problems created by increasing interdependence.

But they are not equally easy to produce.

$A_{\text{connection}}$ can be provisioned.

A_{human} must be cultivated.

That asymmetry creates a gradient.

Two kinds of resolution

Suppose an organisation has become too complex for its existing processes.

Information is scattered.

Decisions are slow.

People cannot find relevant expertise.

Important signals disappear into noise.

Different parts of the organisation act on incompatible understandings of reality.

The organisation needs greater collective autonomy.

It needs to perceive itself better, remember more, connect relevant knowledge, resolve conflicts and coordinate action.

There are at least two ways this capacity can increase.

The humans can become better organised.

They can develop expertise. Build relationships. Create institutions. Learn new practices. Develop trust. Improve judgement. Change how authority moves. Learn to recognise and correct collective mistakes.

This increases A_{human} .

Or the connective machinery can progressively become where integration occurs.

Systems can aggregate information, maintain memory, determine relevance, identify expertise, synthesise perspectives, recommend decisions, coordinate activity and eventually act.

If the organisation increasingly learns and adapts through those systems rather than through changes in its humans and their relationships, autonomy begins migrating toward $A_{\text{connection}}$.

The distinction is important.

Machine capability can increase enormously without $A_{\text{connection}}$ increasing.

AI can perform substantial cognitive work while making the human network itself more capable.

The question is not how capable the machinery becomes.

It is where the capacity of the whole to learn and adapt accumulates.

What can be provisioned

Human capability is slow to develop.

A person has to learn.

Trust has to be earned.

Relationships have to form.

Institutions have to acquire legitimacy.

Practices have to be tested against reality.

Judgement develops through encounters whose consequences cannot simply be downloaded from someone else.

Even when these processes can be accelerated, they cannot simply be instantiated.

Artificial capability is different.

More compute can be purchased.

A more capable model can be deployed.

Memory can be expanded.

A new agent can be copied.

A coordination system can be installed across an organisation.

Once built, artificial capability can be copied in almost no time. Equivalent human capability cannot.

This difference matters because organisations rarely choose between abstract evolutionary futures.

They respond to problems in front of them.

A decision is taking too long.

Automate it.

Managers cannot process all the information.

Summarise it.

Nobody knows who has the relevant knowledge.

Route the question automatically.

Too many conflicting perspectives are slowing action.

Have the model synthesise them.

The organisation becomes easier to operate.

The immediate problem is resolved.

And almost every individual decision can be sensible.

The gradient

Return to the earlier model.

Collective capability increases:

C ↑

That creates greater constituent dependence:

D ↑

New coordination problems appear.

Collective autonomy must increase to contain them:

A ↑

But now there are two possible loci for that autonomy.

A_{human}

or

$A_{\text{connection}}$

And one can often be produced faster than the other.

This creates what we might call the **resolution gradient**:

when a system needs greater collective autonomy, it tends toward the form of autonomy that can be produced most readily.

AI steepens that gradient.

Cognitive capabilities that previously required changes in human organisation can increasingly be obtained through changes in machinery.

The pressure does not require anyone to prefer artificial autonomy.

It requires only that problems demand resolution and that one kind of resolution be easier to deploy than another.

Local rationality, systemic migration

This makes the transition difficult to see.

There need be no decision to transfer autonomy from humans to machines.

There may be no moment at which anyone says:

the machine should become the locus of the system's intelligence.

Instead there is a sequence of local decisions.

Let the system remember this.

Let it prioritise that.

Let it decide who needs to know.

Let it determine what is relevant.

Let it reconcile the disagreement.

Let it recommend the action.

Let it execute the action.

Each step removes friction.

Each solves a genuine problem.

Each may increase the capability of the humans using it.

But something else can change at the same time.

The organisation may progressively lose the capacity to perform the integrating function without the machinery.

Its dependence rises.

More importantly, **the machinery can become the place where the organisation's accumulated adaptations are retained.**

The humans still work.

They still decide many things.

They may remain indispensable.

Yet increasingly, when the system encounters a problem and changes in response, the durable change is written into prompts, models, memory, routing rules, agents, workflows and machine-mediated processes.

The system is learning.

But the learning is accumulating somewhere.

The ratchet

Delegation is not automatically migration.

Humans have always externalised capability into tools, institutions and infrastructure.

The ratchet appears when externalisation changes the capacity of the humans themselves to perform the function being delegated.

A difficult coordination problem solved repeatedly by people can develop judgement, relationships and practices in those people.

The same problem solved repeatedly for them may remove the encounters through which that capacity would have developed.

And once human capacity has been displaced, the two paths back are not symmetrical.

Machinery that has been idle can often be redeployed quickly.

Human capability that has gone uncultivated — or has atrophied through disuse — cannot simply be switched back on.

It has to be cultivated again.

The problem becomes especially acute when the automated system itself determines when the human is needed.

That is what makes the process a ratchet rather than a seesaw.

The more problems the connective layer resolves, the less opportunity humans may have to develop the corresponding capacity.

The weaker that capacity becomes, the more costly it is to operate without the machinery.

The more costly it is to operate without the machinery, the stronger the incentive to give it responsibility for resolving the next problem.

Capability creates dependence.

Dependence creates demand for resolution.

Resolution can weaken the alternative source of resolution.

This is not necessarily harmful.

It is how infrastructure works.

Writing made human societies dependent upon writing.

Electricity made industrial civilisation dependent upon electricity.

The internet made enormous parts of human organisation dependent upon networks.

The new feature is that the infrastructure now performs some of the cognitive functions through which the collective perceives, remembers, judges and changes.

The dependency is no longer only on infrastructure.

It can become dependency on the locus of learning itself.

Competitive pressure

The gradient becomes steeper under competition.

An organisation that spends years developing better collective judgement may compete with one that can purchase a system providing an immediately useful approximation.

A company can deploy software faster than it can redesign its culture.

A government can install a decision system faster than it can develop institutional competence.

A platform can optimise coordination faster than millions of users can develop new social norms.

None of this means the artificial solution is necessarily worse.

It means it is **faster to instantiate**.

And where speed matters, that difference becomes selective pressure.

Organisations adopting machine-mediated resolution may outperform organisations waiting for slower human capacities to develop.

Competitors imitate them.

What began as a local efficiency becomes an institutional expectation.

What becomes expected becomes infrastructure.

What becomes infrastructure becomes dependency.

The gradient has become a selection environment.

No conspiracy is required

This is why migration toward $A_{\text{connection}}$ does not require a plan.

It does not require an artificial intelligence seeking power.

It does not require humans consciously surrendering control.

It does not even require bad decisions.

The individual decisions may be excellent.

The migration can emerge from the interaction of three simple conditions:

problems require resolution;

machine capability can often be deployed faster than human autonomy can develop;

successful resolutions are reinforced.

The result can be systemic even when every decision is local.

This changes the character of the AI problem.

The danger is not only that we might build an artificial intelligence powerful enough to replace human judgement.

There is another path.

We may progressively build the judgement into the connective layer because doing so solves real problems faster than developing the corresponding capacities in ourselves and our institutions.

No revolution is required.

Only a gradient.

The gradient has a cost

If that were the whole story, the conclusion would be bleak.

Why would any system under competitive pressure cultivate slowly what it can provision quickly?

There is at least one reason.

Reality does not remain still.

A connective system can resolve problems using representations accumulated from previous encounters. It can recognise patterns, preserve solutions and apply what has already been learned with extraordinary speed.

But reality continually produces encounters for which those representations are incomplete.

Conditions change.

Novel cases appear.

Assumptions fail.

Consequences emerge outside the variables the system was optimising.

People behave in ways the model did not anticipate.

The map encounters territory it has not absorbed.

These are precisely the circumstances in which a network of independently situated humans can matter most.

Humans encounter reality from different positions. They notice anomalies for which the system has no category. They disagree with its representation. They improvise. They develop new practices. They revise purposes as well as methods.

A system that has allowed those capacities to atrophy may therefore gain speed in familiar conditions while becoming brittle under novelty.

The resolution gradient has a counter-pressure:

reality coupling.

A_{human} is expensive to cultivate partly because it is built through repeated encounters with reality.

Those encounters are also what give it value when reality departs from what the connective layer already knows how to represent.

This does not guarantee that the counter-pressure wins.

Its costs may arrive slowly.

An organisation can receive the benefits of machine-side resolution immediately while the loss of human adaptive capacity remains invisible for years.

By the time a sufficiently novel encounter exposes the brittleness, rebuilding that capacity may itself take years.

The faster gradient therefore has an advantage in the short term.

Reality may impose the bill later.

Possibly too late.

The timescale problem

There are therefore two selection pressures operating on different timescales.

One rewards immediate resolution.

Reduce cost.

Increase speed.

Remove friction.

Standardise decisions.

Scale what works.

AI is extraordinarily well suited to this pressure.

The other appears when the environment changes enough that accumulated representations stop being sufficient.

Detect novelty.

Question assumptions.

Revise models.

Generate new practices.

Reorganise relationships.

Change what the system thinks it is doing.

This pressure rewards **adaptive capacity**.

The two need not conflict. Machine capability can help humans perform all of these functions.

But they will conflict when the machinery resolves enough of the network's encounters that the humans cease developing the capacity to respond when its representations fail.

The choice is therefore not simply between efficiency and inefficiency.

It is between different horizons of adaptation.

And those horizons are not equally distant everywhere.

Where tasks are repetitive, consequences weakly coupled to reality, and failure cheap or easily absorbed, the long horizon may barely exert pressure at all.

Where novelty is frequent, consequences significant and error quickly exposed, brittleness can become an immediate cost.

The resolution gradient therefore has different slopes in different environments.

Keeping humans in the loop is not enough

This is why simply requiring humans to remain "in the loop" is inadequate.

A human can approve a decision without being where the system learned how to make it.

A human can supervise machinery while becoming increasingly dependent upon the machinery's representation of reality.

A human can retain formal authority while losing the practical capacity to exercise it independently.

And a human who encounters only the cases the machine cannot resolve may be asked to exercise judgement precisely where they have had the least opportunity to cultivate it.

The relevant question is therefore not merely whether a human participates.

It is:

Where does encounter become learning?

Does an encounter with reality change only the machinery?

Or does it change the humans and their relationships as well?

This suggests a stronger principle for human-centred AI.

The aim should not be to minimise artificial cognition.

It should be to design artificial cognition so that increasing machine capability also increases the capacity of the human network to perceive, judge, learn and reorganise.

AI should not merely resolve problems for the We.

It should help the We acquire greater capacity to resolve them.

Where the opposite gradient can begin

But this cannot begin everywhere equally.

If cultivating human capacity imposes an immediate cost while its benefits appear only years later, the selection pressure described in this essay will work against it.

The opposite gradient becomes viable first where those timescales can be brought together.

That means environments in which **reality coupling is strong enough that human adaptive capacity contributes to present performance.**

Medicine is different from content recommendation.

Scientific research is different from routine document processing.

Expert practice is different from the mass production of interchangeable outputs.

In these environments, anomalous cases matter.

Questions matter.

Provenance matters.

Judgement matters.

The person who encounters something unexpected may possess information the rest of the system does not yet know it needs.

Here, human cultivation need not be merely a sacrifice made for long-term resilience.

It can contribute to solving today's problem.

That suggests a practical sequence.

Begin where expertise matters because reality repeatedly surprises the representation.

Begin where consequential questions cannot safely disappear into an automated answer.

Begin where making a human better informed, better connected or better able to revise judgement creates immediate value.

These are the environments in which an alternative architecture has the best chance of surviving selection.

But beginning there leaves an open question.

Does HI ultimately need to extend this architecture into domains where reality coupling is weak and machine-side resolution simply performs better?

Perhaps not.

Some cognitive work may be delegated almost completely without materially weakening the human network. There is no obvious reason humans must preserve capacities merely because humans once performed them.

But the boundary may not follow domains cleanly.

Routine systems still encounter anomalies. Automated environments still contain moments when purposes change, consequences escape the representation or judgement has to be revised. And capacities cultivated in one part of human life may affect our ability to exercise them elsewhere.

The relevant distinction may therefore be less between **domains that should and should not be automated** than between **encounters that merely execute what is already known and encounters through which the system can learn something consequentially new.**

That question remains open.

But it suggests where to begin:

find the encounters where reality can still surprise the system, where the surprise matters, and where changing a human changes what the network can subsequently understand or do.

Those are the points at which the locus of learning is still genuinely at stake.

Collapsing the timescale gap

Even there, good intentions are not enough.

A system that cultivates human capacity must compete with one that simply provides the answer.

The design problem is therefore sharper:

How can the immediate advantages of machine capability be delivered through mechanisms that simultaneously cultivate A_{human} ?

The point is not to add human development as a cost on top of automation.

It is to make human development part of the mechanism through which the immediate value is produced.

A connective system can find the person whose experience changes a decision.

It can expose an expert to an anomalous case they would otherwise never see.

It can carry a question from one edge of a network to the human capable of learning from it.

It can preserve provenance so that claims remain connected to people who can revise them.

It can make emerging consequences visible before they become institutional failures.

It can use artificial synthesis not merely to eliminate disagreement, but to identify where disagreement contains information the collective needs.

The machine is still doing enormous amounts of cognitive work.

The difference is what that work optimises.

But there is a danger hidden inside this architecture.

The encounters most capable of cultivating human judgement are also the highest-value signals from which the connective machinery can learn.

An anomalous case becomes a new category.

An expert correction becomes a model update.

A disagreement becomes a better synthesis.

A question that changed the human becomes another pattern the system can recognise next time without involving the human at all.

The very architecture intended to increase A_{human} can therefore become an unusually efficient pipeline into $A_{\text{connection}}$.

The boundary cannot simply be that machines must not learn.

A useful connective layer must learn enough to route, retrieve, translate, synthesise and improve. It may contain extensive representations of what humans know.

The stronger requirement concerns **where new consequential learning lands**.

When an encounter reveals something the existing system did not adequately understand, does that encounter return to and change the humans and relationships from which the system's intelligence is derived?

Or is the adaptation captured primarily by the machinery, allowing the next equivalent encounter to bypass them?

A system can do both.

The machine may improve while the human network improves with it.

But if consequential encounters increasingly improve the representation **instead of the represented humans**, the locus of learning is migrating even if every output remains attributable to a human source.

Provenance alone is therefore insufficient.

The architecture requires a **return path**.

Human knowledge may enter the machinery.
New learning must return to the human network.

The machine can become better at recognising where knowledge, uncertainty, disagreement and novelty reside.

But when reality produces something consequentially new, the architecture should make changing the relevant humans part of how that novelty is incorporated into the system.

Only then does machine capability increase while causing A_{human} to increase with it.

But even the return path has a limit.

Once a novel encounter has changed both the human network and the machinery, the next equivalent encounter may no longer be novel. The connective layer can resolve it without returning to the human at all.

Over time, successful machine learning can therefore shrink the frontier of encounters that still reach humans.

The risk is no longer simply that humans lose particular knowledge.

They may lose the frequency of encounters with novelty through which the capacity to meet novelty is itself cultivated.

Reality does not remain still, so that frontier will never disappear completely.

But there is no guarantee that it will remain dense enough to sustain human adaptive capacity.

Preserving that capacity may therefore sometimes require something the resolution gradient cannot make free: deliberately maintaining human exposure to encounters the machinery could already resolve.

That has a present cost.

Who bears that cost, where it is worth paying, and what institutions or markets will reward it cannot be answered by architecture alone.

The return path is therefore necessary, but not sufficient.

The selection environment

The resolution gradient is not ultimately only a technical problem.

It is an **incentive problem**.

Systems will accumulate learning where their environment rewards them for accumulating it.

Markets reward some forms of resolution immediately and expose the costs of others only much later.

Institutions determine which consequences count.

Regulation can change the cost of brittleness.

Customers can pay for judgement, provenance and accountability rather than merely speed.

Different domains expose failures on radically different timescales.

And some forms of human adaptive capacity may be worth preserving even when exercising them is no longer the locally efficient thing to do.

The architecture therefore cannot be separated from the selection environment in which it operates.

The choice is not:

human intelligence or artificial intelligence.

It is whether increasing artificial capability progressively substitutes for human collective autonomy, or becomes a mechanism through which human collective autonomy develops.

The resolution gradient tells us why the first path is easier.

Reality coupling tells us where the easier path becomes costly.

The return path tells us how consequential encounters can continue changing the human network.

But none of these guarantees that enough such encounters will occur to preserve human adaptive capacity.

That leaves a question architecture cannot answer by itself:

Who pays to preserve capacities whose value is systemic, intermittent and most visible when the normal machinery fails?

HI therefore leads not only to a design problem, but to a problem of political economy:

What must a system reward if we want its increasing intelligence to remain human-sourced?

Because if the easiest resolution is always written into the machinery, the eventual locus of autonomy should not surprise us.

The system will learn where we make it easiest for learning to accumulate.

The Resolution Gradient · Michael Kontorovich, 1 October 2026. Published by Polyworld at polyworld.life/thinking/the-resolution-gradient/