GOT-V1-3b-4

# On the Origin of Responsibility

Primitive Causal Attribution, Responsibility Provenance, and Warrant Under Finite Conditions

---

## Abstract

Responsibility is not first blame, guilt, duty, intention, legal liability, moral desert, institutional role, social expectation, or conscious agency. These are later enrichments, allocations, and warrant regimes around a more primitive relation. At its origin, responsibility is causal and attributional: a consequential downstream branch becomes dependent on a discriminable distinction because the organisation of a specified locus makes that distinction operative.

Causation alone is too broad. A stone can break glass, a storm can redirect a river, a wall can block movement, and a passive constraint can alter trajectory. If responsibility is attributed wherever causal contribution appears, responsibility collapses into causation and loses discriminatory value. Non-Trivial Computation supplies the inherited causal discriminator: an organised pathway makes downstream outcome depend counterfactually on a discriminable distinction under specified reference and comparator conditions. The preceding paper then identifies Power as the indexed potential of an Information Processing Locus to alter downstream branch-realisation through such distinction-conditioned organisation. Responsibility begins one step later. It asks which consequential application, non-application, transformation, maintenance, or field-conditioning effect of that potential is attributable to the locus under the actual conditions of the relation. The relevant consequence may be a realised branch, but it may also be an altered condition of branch availability where one locus's distinction-conditioned organisation changes what another locus can physically realise without itself producing Counterfactual Exclusion.

This paper defines primitive responsibility as locus-indexed attribution of consequential branch-relation dependence to distinction-conditioned organisation. It then develops Responsibility Provenance: the recoverable path by which one or more local transformations condition a downstream consequence. Provenance may be immediate, sequential, nested, distributed, retained, or recursively abstracted. It does not simply transfer between loci. A downstream transformation does not erase upstream provenance, an upstream contribution does not absorb every later transformation, and causal proximity does not determine responsibility relevance.

As the architecture becomes richer, responsibility also becomes richer. A locus may discriminate more alternatives, retain distinctions across longer horizons, abstract and recombine them, preserve residual, represent counterfactual branches, detect mismatch, revise, stop, or reroute. These capacities do not create responsibility by complexity alone and do not produce a moral score. They define a Responsibility Gradient: an ordinal, relational increase in the depth and breadth of distinction-conditioned causal contribution that can be attributed under the locus's own operative frame.

The paper then separates Responsibility Provenance from Responsibility Warrant. A social, moral, legal, organisational, or institutional regime may allocate answerability, duty, burden, liability, or decision ownership to a locus. Such allocation may be necessary and legitimate. It

does not thereby establish the causal provenance of the consequence. Responsibility distortion begins where an abstracted warrant, role, label, or convenient accountable locus is allowed to carry provenance-strength beyond the relation it preserves.

The governing rule is simple: responsibility must not carry attribution-strength beyond the recoverable distinction-conditioned causal contribution of the locus under the reference, comparator, capacity, and temporal conditions being claimed. Where responsibility is assigned by warrant, the warrant must not be treated as provenance merely because authority, role, convention, procedure, or institution supplies it.

Responsibility therefore precedes trust in the causal sequence. Consequential transformations can be attributable before one locus relies on another. Trust becomes necessary later because finite loci cannot continuously reconstruct the responsibility provenance of every inherited relation upon which further action depends. Responsibility identifies where consequential transformation is attributable. Trust explains how finite systems continue across such dependency without reopening the entire path at every use.

---

# 1. Introduction: Responsibility before agency

This paper concerns responsibility as a primitive causal-attribution relation under finite conditions.

It does not offer a complete theory of moral responsibility, blame, guilt, culpability, freedom, desert, negligence, legal liability, political accountability, parental duty, professional obligation, institutional governance, or punishment. Those domains matter. Human responsibility is saturated with intention, expectation, duty, authority, vulnerability, power, social role, retrospective judgement, and moral evaluation. A person may be praised, blamed, sanctioned, excused, compensated, required to repair, or held answerable. An institution may assign a duty to one role, liability to another, and operational authority to a third. None of those practices is dismissed here. The question is whether they are the primitive condition from which responsibility originates.

The primitive question is earlier.

Under what conditions can a consequential change in system state be attributed to the distinction-conditioned organisation of a specified locus?

That question appears before law, before explicit morality, before language, and before reflective agency. It appears wherever an organised system can discriminate among alternatives and where later trajectory depends on how that distinction is carried. A temperature-sensitive switch, a bacterium responding to a chemical gradient, a repair pathway responding to molecular damage, a nervous system selecting among action branches, a person applying a retained category, a machine classifier routing a case, and an institution applying a rule may all occupy different positions in such an architecture. They do not thereby become equivalent. Their capacities, horizons, abstractions, residuals, available alternatives, and downstream leverage differ radically. But the attributional question can be asked across them without first importing the human moral vocabulary by which responsibility is usually recognised.

The need for a primitive account arises because ordinary responsibility is usually approached too late. We begin with the adult person, the legal actor, the office-holder, the parent, the professional, the corporation, the state, or the machine operator. The concept arrives already carrying a theory of agency. It may presume that responsibility belongs to those who know,

intend, choose, control, consent, foresee, occupy a role, or can be sanctioned. Those may be important discriminator conditions for later forms of responsibility. They cannot establish the primitive if a thinner attributional relation already exists below them.

The inherited architecture supplies reason to expect such a relation. Physical difference may become operational distinction for a finite locus. Distinction may be registered, retained, and redeployed. Applied distinction may change downstream branch-realisation. Non-Trivial Computation disciplines the stronger causal claim: the system's own organisation must make a discriminable condition matter to what happens next, and that dependence must survive meaningful comparison against a condition in which the proposed mediating organisation is removed, disabled, bypassed, scrambled, or otherwise prevented from carrying the distinction.

The preceding paper isolates the power consequence of that architecture. Power is the indexed potential of an Information Processing Locus to alter downstream branch-realisation through organisation-mediated discriminable distinctions under stated conditions. It separates potential from application and consequence, fixes the relevant Counterfactual Inclusion Field to the IPL, and identifies Counterfactual Exclusion where applied distinction produces a branch-realisation event.

Responsibility does not redefine that power relation. It inherits the IPL, the distinction field, the Counterfactual Inclusion Field, comparator pressure, the distinction between potential and application, and Counterfactual Exclusion where application has produced a realised branch. But Responsibility need not wait for Counterfactual Exclusion. A locus's power-bearing organisation may also alter the accessibility, probability, persistence, reachability, timing, or later branching conditions of another locus's Counterfactual Inclusion Field while the relevant potential remains unapplied. Responsibility asks whether that consequential branch relation — realised or field-conditioning — is attributable to the locus under the actual conditions of the relation.

NTC therefore remains the causal discriminator beneath both papers. Power asks what the locus could make consequential. Responsibility asks what consequential application or non-application of that potential is attributable under the event's actual conditions.

The difference is small at primitive scale and decisive at distributed scale. In a simple switch the organisation, distinction, transformation, and consequence may lie close together. In a nested system an upstream locus may classify, another may retain, another may compress, another may transmit, another may trust, another may formalise, and another may finally act. The final physical movement may be the most proximal cause while an earlier abstraction remains a major determinant of the available branch. Conversely, an upstream classification cannot be assigned responsibility for everything downstream merely because later loci encountered it. Every local transformation must be distinguished.

Responsibility is therefore not a substance that moves through a causal chain. It is not a token that can be passed from one actor to another. It is not a possession attaching globally to a person, system, role, or institution. It is an indexed attribution relation. The relevant question is always responsibility for what transformation, through which distinction-conditioned organisation, under which reference conditions, relative to which comparator, across what horizon, and with what available alternatives and residual.

This indexing becomes more important as abstraction deepens. A finite locus never acts from total state. It encounters relation through boundary, interface, scale, discrimination regime, abstraction, registration, retention, redeployment, comparator pressure, closure demand, and residual limit. Its responsibility cannot therefore be measured against an exterior omniscient reconstruction as though every later-discovered distinction had been available within the locus

at the time. Nor can limited access become a universal excuse. Where a locus could discriminate a mismatch, retain residual, reopen a closure, compare alternatives, update, stop, or reroute, those capacities are part of the responsibility relation.

The paper develops this structure in stages. It first separates causation from responsibility. It then derives primitive responsibility from the inherited NTC discriminator through the power relation established in the preceding paper, without redefining NTC, IPL, or Power. It decomposes the responsibility relation into locus, operative distinction, mediating organisation, branch consequence, comparator, and reference conditions. It defines Responsibility Provenance as the recoverable path of responsibility-bearing transformation and shows how such provenance becomes sequential, nested, and distributed without becoming undifferentiated. It then develops responsibility-relevant capacity and the Responsibility Gradient, followed by the Immediate-Frame Principle and the temporal discipline required for retrospective attribution.

Only after the primitive and distributed causal architecture has been established does the paper introduce Responsibility Warrant. Human systems often need to allocate duty, answerability, liability, supervision, stewardship, or decision ownership even where causal provenance is distributed or only partially recoverable. Such warrants may be necessary Soft Horizons for action. The failure begins when the warrant is reified as if it were the causal provenance itself.

The paper ends at the point where responsibility hands to trust. Once consequential transformations are distributed across finite loci, no receiving locus can reconstruct every upstream discrimination, retained abstraction, residual, comparator, and transformation before every use. Trust becomes the reliance mechanism by which some of that work is allowed to remain backgrounded. But trust cannot create, transfer, or erase responsibility provenance. It operates over a causal architecture that responsibility has already made attributable.

The governing rule is therefore:

Responsibility must not carry attribution-strength beyond the recoverable distinction-conditioned causal contribution of the locus under the reference, comparator, capacity, and temporal conditions being claimed.

Where responsibility is assigned by social, moral, legal, organisational, procedural, or institutional warrant, the warrant must not be treated as responsibility provenance merely because an authority, role, convention, procedure, or institution supplies it.

Responsibility is not the moral judgement placed on causation.

It is the attributional structure that makes later judgement possible.

---

# 2. Causation is too broad

The first mistake is to make responsibility coextensive with cause.

Causes are everywhere. A stone striking glass changes the later state of the glass. A storm changes pressure, flow, temperature, erosion, and structural load. A wall blocks movement. A narrow channel redirects fluid. A chemical concentration alters reaction rate. A missing component changes what a mechanism can do. A passive constraint can be indispensable to an outcome. None of this is trivial in the ordinary sense. Causal structure is the physical condition from which every later responsibility claim must be made. But if every causal contribution is already responsibility, the concept does no additional work.

Responsibility requires a discriminator.

The discriminator cannot be consequence magnitude. A storm may produce catastrophic damage while a thermostat produces only a local switching event. Magnitude tells us something about the scale of consequence, not whether the causal path contains responsibility in the technical sense being developed here.

The discriminator cannot be complexity. A weather system may be vastly more complex than a simple regulatory circuit. Complexity may arise from many interacting variables without there being a local organisation that treats one discriminable condition differently from another in a way that routes a later branch. A responsibility account built on generic complexity would therefore reproduce the same failure as unrestricted computational redescription: the category would expand faster than its explanatory power.

The discriminator cannot be causal proximity. The final physical movement before an outcome may be a finger pressing a control, a valve opening, a document being filed, a machine executing an instruction, or a structural element failing. Proximity identifies one location in the path. It does not tell us where the distinctions that conditioned the branch were formed, retained, transformed, excluded, or made authoritative.

The discriminator cannot be external describability. An analyst can classify almost any physical event into alternatives and reconstruct a causal path after the fact. The analyst may say that the stone was on one side rather than another, that the storm crossed one pressure threshold rather than another, or that a chemical state fell into one class rather than another. Those distinctions may be useful to the analyst. Responsibility requires that the relevant distinction be operative for the evaluated locus rather than merely available to an exterior model.

The discriminator cannot be usefulness, harm, success, or failure. A system may produce an advantageous outcome through passive structure. It may produce harm through an accidental physical interaction. It may fail despite disciplined discrimination. It may succeed despite a weak or accidental path. Outcome valence and outcome success therefore belong to later evaluation. They do not define the primitive responsibility relation.

This leaves a narrower possibility. Some causal paths are not merely physical sequences in which prior state produces later state. They contain an organisation that distinguishes alternatives and makes what happens next depend upon which alternative is operative.

A doorstop changes whether a door closes. It is a passive constraint. The downstream state differs because the physical situation differs, but the doorstop does not contain an organised discriminator that registers an alternative and routes a later operation through that distinction.

A thermostat marks a stronger relation. Temperature relative to a set point becomes an operative distinction. The device is organised so that above-threshold and below-threshold conditions route later states differently. Disable or bypass the sensing-and-switching pathway and the downstream branch no longer tracks the temperature distinction in the same way. The device does not become morally responsible. It does not become a person. It does not become alive. But a new causal structure is present: an organised locus makes a discriminable difference matter to what happens next.

That is the threshold responsibility inherits.

The stone and the thermostat are both causal. Only the thermostat supplies the relevant distinction-conditioned causal attribution under the stated relation.

This does not mean every thermostat event is automatically a responsibility event. Reference conditions still matter. If the sensor is disconnected, the threshold is not operative. If the output

is fixed regardless of temperature, the apparent distinction does not condition the branch. If an external observer merely labels two temperatures while the device is insensitive to them, the distinction remains exterior to the locus. If the comparator is weak, the attribution remains weak.

The first boundary can therefore be stated precisely:

Causal contribution is necessary for responsibility attribution, but causal contribution alone is insufficient.

Responsibility begins only when causal consequence can be localised to organisation-mediated distinction-use under conditions that make the attribution testable.

This is why the origin of responsibility is not found by adding morality to cause. It is found by identifying the point at which cause becomes attributable through operative discrimination.[1]

---

# 3. From Power to primitive Responsibility

Non-Trivial Computation remains the causal discriminator inherited beneath the present argument. It establishes that an organised distinction makes downstream outcome depend on an operative alternative under comparator pressure. The preceding paper then asks what potential exists because that difference can be made: Power.

This paper does not redefine NTC, IPL, or Power. It inherits from the power relation the localised IPL, the available distinction field, the Counterfactual Inclusion Field, comparator pressure, the distinction between potential, application, and consequence, and Counterfactual Exclusion where distinction-conditioned application has produced a branch-realisation event.

Power asks: what branch-changing potential is physically included for the IPL under the indexed conditions?

Responsibility asks: which consequential relation is attributable when that potential is applied, left unapplied, retained, maintained, governed, or allowed to condition the possibilities available to another locus under the actual conditions?

At primitive scale, the two questions may nearly coincide. If a thermostat's threshold architecture is the organisation whose integrity is required for temperature-dependent switching, the downstream switching branch is attributable to that locus under the relevant reference conditions. The responsibility claim is not that the thermostat deserves praise or blame. It is that a particular consequential transformation belongs to its distinction-conditioned causal operation rather than to the external analyst's description or to passive background causation alone.

Primitive Responsibility can therefore be defined as follows:

Primitive Responsibility is the minimal responsibility status earned where a consequential downstream branch relation — including branch-realisation or a material condition of branch availability — is counterfactually dependent upon operative or retained distinction-conditioned organisation localisable to a specified locus under stated reference and comparator conditions.

Each part of the definition matters.

The responsibility is primitive because no richer agency condition has yet been established. The locus need not possess language, reflective self-model, moral understanding, intention, or capacity to justify its operation.

The responsibility is consequential because a downstream branch relation must differ. Consequential does not mean morally important, and it does not require that a terminal branch has already been realised. The difference may concern realised state, probability, timing, persistence, accessibility, reachability, or the conditions under which later branching remains possible.

The dependence is counterfactual because mere correlation or coexistence is insufficient. The question is whether the relevant downstream branch relation would remain materially the same if the proposed distinction-conditioned pathway, retained organisation, or field-conditioning condition were absent, disabled, altered, withdrawn, or otherwise prevented from doing the claimed work while enough relevant context remained stable for comparison.

The distinction must be operative because an exterior label does not become a system operation simply by being analytically convenient. The system must be physically organised so that the relevant alternative enters its own downstream path.

The mediation must be localisable because responsibility without a locus becomes floating attribution. The evaluated locus may be simple or distributed, but the analysis must identify the organisation whose transformation is being claimed.

The conditions must be stated because the same physical carrier may participate in different relations under different boundaries, scales, noise regimes, horizons, and comparators.

This definition establishes a low threshold on purpose. A primitive should not be forced to inherit all of its later enrichments. If moral intention were built into the origin, sub-human and sub-linguistic responsibility could never be examined. If legal answerability were built into the origin, responsibility would become dependent on institutions that themselves require explanation. If conscious foresight were built into the origin, large regions of living and machine causation would disappear from the concept by stipulation rather than discriminator.

The low threshold does not flatten the later differences. A thermostat's primitive responsibility is extremely narrow. It is indexed to a small discrimination set, a limited temporal horizon, a fixed architecture, a small branch-space, little or no endogenous representation of residual, and limited ability to revise its own threshold conditions. A human decision-maker may discriminate many alternatives, retain history, represent residual, form abstractions, anticipate downstream effects, revise categories, stop an action, ask for further evidence, and inspect the provenance of a trusted carrier. An institution may distribute those operations across many loci. These differences matter. They are not reasons to deny the primitive. They are reasons to develop the gradient later.

The NTC root also prevents a common mistake: attributing responsibility to the widest available frame. A chemical environment may be necessary for a cell's response. A road network may be necessary for a driver's route. A training distribution may be necessary for a model output. A legal system may be necessary for a court procedure. Necessity in the broad causal background does not tell us which organisation performed the distinction-conditioned transformation under review.

Responsibility therefore does not attach to everything that made the outcome possible. It attaches, in the primitive technical sense, to the localisable operations through which discriminable alternatives became consequential.

This is also why responsibility is not identical to causal leverage in the broadest sense. A large passive force may have more effect on an outcome than a small discriminating system. Responsibility provenance isolates a particular kind of leverage: leverage achieved because organisation made a distinction operative.

The gain is conceptual economy. No new metaphysical property is required. Responsibility does not suddenly appear when a society declares it, when a mind reflects upon itself, or when a moral vocabulary is acquired. The primitive becomes visible at the point where the existing architecture already supports attributable distinction-conditioned consequence.

NTC is therefore not responsibility itself.

It is the inherited causal condition that makes primitive responsibility attribution possible.

---

# 4. The Responsibility relation: locus, distinction, branch, comparator

Responsibility is usually spoken as though it were possessed.

A person is responsible. A parent is responsible. A company is responsible. A machine is responsible. A government is responsible. The grammar is convenient, but it hides the relation. It invites responsibility to attach globally to a named entity rather than to a specified transformation.

The primitive architecture requires a different grammar.

A locus is responsible for a particular consequential transformation insofar as its distinction-conditioned organisation made the relevant downstream branch depend on an operative alternative under stated conditions.

Responsibility is therefore indexed.

At minimum, a technical responsibility claim must recover eight elements: the locus, the operative distinction, the availability condition and immediate frame under which that distinction and the relevant alternatives stood for the locus, the mediating organisation, the consequential branch relation, the comparator, the reference conditions, and the closure demand. Where the attribution remains incomplete, residual and update must remain attached to the claim.

Where the claimed consequence is cross-locus field-conditioning, the affected locus must also be identified, together with the material property or region of its Counterfactual Inclusion Field alleged to have changed. This is a conditional cross-locus field rather than a ninth universal element.

The locus identifies where the transformation occurs. It may be a molecular regulatory pathway, a cell, an organism, a person, a machine subsystem, a human-machine workflow, or a distributed institutional process. The correct locus is not decided by familiarity. It is determined by the organisation whose integrity is required for the distinction-conditioned branch relation being claimed.

The operative distinction identifies what difference mattered. It may concern temperature, molecular form, concentration, timing, spatial position, signal class, retained category, measurement interval, legal classification, risk threshold, or some other discriminable

alternative. The distinction must be available to the locus under the relevant regime. A distinction reconstructed only by an exterior observer does not automatically enter the locus's responsibility frame.

The mediating organisation identifies how the distinction entered consequence. This is the pathway, architecture, retention mechanism, rule, model, threshold, classification process, or other organised relation that caused alternatives to route differently. Without mediation, the analyst may have identified correlation or passive constraint rather than responsibility-bearing transformation.

The consequential branch relation identifies what downstream possibility or realisation changed because the distinction-conditioned organisation was operative. The consequence may be a realised branch, but it may also be a material change in probability, timing, persistence, accessibility, reachability, or later branching conditions. A thermostat may be responsible for switch activation without being responsible for the entire thermal state of a building. An upstream locus may likewise be responsible for materially constraining another locus's available possibilities without becoming responsible for every branch that the downstream locus later realises.

The comparator supplies pressure against attribution inflation. What happens if the relevant distinction-conditioned pathway is disabled, neutralised, bypassed, or replaced while other material context is preserved? If the outcome remains materially unchanged, the responsibility claim weakens. If the branch changes in the predicted way, the attribution strengthens.

The reference conditions state boundary, scale, horizon, noise regime, driving conditions, component inventory, interface, and other assumptions material to the claim. Responsibility is not made arbitrary by indexing. Indexing makes the claim testable. It states the conditions under which the attribution is being made and the conditions under which it should be revised.

The closure demand states what the responsibility attribution is being asked to establish. A path may be sufficient for a narrow primitive transformation while remaining incomplete for distributed provenance, retrospective failure, or a stronger warrant-relevant claim. Responsibility strength must therefore remain indexed to the attribution actually being demanded.

These eight elements are evaluated within the power relation inherited from the preceding paper. A realised branch attributed to the locus must fall within the relevant Counterfactual Inclusion Field under the stated conditions. Where application is claimed, the audit should recover the relevant Counterfactual Exclusion relation. Where field-conditioning is claimed without application, the audit should instead recover the comparator-stable difference made to the relevant Counterfactual Inclusion Field. Responsibility does not require a separate scalar measure of power, but attribution cannot exceed the branch-changing potential or field-conditioning relation that can actually be recovered.

These eight elements prevent several familiar errors.

They prevent container collapse. A process occurring inside an entity does not automatically belong to the entity as a whole. The relevant locus may be narrower, wider, or differently distributed.

They prevent outcome collapse. A bad outcome does not prove irresponsible transformation, and a good outcome does not prove responsible transformation. The pathway must be examined.

They prevent role collapse. Occupying a named position does not establish which distinction-conditioned transformations the role actually performed.

They prevent hindsight collapse. A later observer may reconstruct distinctions and causal branches that were not available within the earlier locus's operational frame.

They prevent external-interpreter substitution. An analyst's ability to describe a state as one class rather than another does not establish that the evaluated locus used that distinction.

The indexed grammar also explains why the same locus may carry radically different responsibility across different relations. A thermostat may carry primitive responsibility for threshold-conditioned switching but none for its own manufacture. A clinician may carry responsibility for a treatment decision under available evidence but not for an unknown molecular pathway outside present access. A model developer may carry responsibility for a training or deployment transformation but not automatically for every downstream use performed by independent loci. A parent may carry responsibility provenance for some educational conditions while a social warrant assigns broader duties that do not map one-to-one onto causal production of the child's educational state.

The sentence "X is responsible" is therefore technically incomplete.

The stronger form is:

For relation R, locus L bears responsibility attribution for consequential branch relation B insofar as operative distinction D was mediated by organisation O under reference conditions C and survives comparator pressure K.

No formal notation is required for ordinary use. The point is architectural. Responsibility must remain attached to the transformation that earns it.

Responsibility is not a character trait of a locus.

It is a relation between a locus, a distinction-conditioned transformation, and a consequence.

---

# 5. Responsibility Provenance

Once responsibility is indexed to a transformation, the next problem is provenance.

A consequence rarely arrives from one isolated operation. Even simple systems sit inside causal fields. More complex systems receive prior distinctions, inherit retained states, transform earlier outputs, encounter environmental constraints, and produce new relations that later loci may themselves use. A responsibility claim therefore requires more than the identification of one proximal event. It requires a recoverable path.

Responsibility Provenance is the recoverable locus-indexed path by which distinction-conditioned organisation contributes to a consequential downstream branch relation under specified reference and comparator conditions.[2]

The word recoverable is important. Provenance does not require omniscient reconstruction of total physical history. That burden would make finite audit impossible. It requires enough recovery to support the attribution being made. The stronger and more consequential the claim, the greater the burden of recovering the transformations material to it.

The word path is equally important. Responsibility provenance is not a list of responsible entities. It is a sequence or network of transformations.

A minimal path may look like this in conceptual form: a distinction becomes available; the locus discriminates or classes it; the distinction is registered or retained; organisation

transforms or deploys it; a downstream branch differs; the output may then become available to another locus.

At primitive scale the path may be short. Temperature crosses a threshold, the thermostat registers the condition, the switching architecture changes state, and the heating branch changes.

At biological scale the path may include receptor binding, signal transduction, retained regulatory state, gene expression, transport, movement, repair, or metabolic redirection. Multiple distinction-conditioned operations can participate in one continuing regime.

At human scale the path may include observation, memory, abstraction, language, judgement, anticipated consequence, action selection, and correction.

At institutional scale the path may include a classification created by one group, encoded in a rule by another, entered into a record, implemented in software, interpreted by an operator, applied to a case, reviewed by a supervisor, and enforced through a later decision. The final action may be physically local while the conditions that shaped its branch have accumulated across many earlier transformations.

Responsibility provenance therefore differs from causal provenance in scope.

Causal provenance is the wider physical history relevant to the outcome. It may include passive constraints, background conditions, environmental forces, component failures, energy supply, material structure, and many other causes.

Responsibility provenance follows the subset of the causal architecture in which distinction-conditioned organisation did consequential work for the evaluated relation.

This does not make passive causes unimportant. A bridge may fail because material fatigue exceeded tolerance. A storm may supply the load. A maintenance classification may have been ignored. A sensor may have failed. A decision rule may have deferred inspection. The causal explanation may require all of these. Responsibility provenance asks which parts involved organised discrimination that conditioned the consequential branch and how those transformations relate.

The distinction also prevents responsibility from becoming a substitute for causal science. Responsibility analysis should not overwrite physical causation with agency language. It depends on adequate causal understanding. Where the physical relation is unresolved, the responsibility claim inherits that residual.

Provenance also has direction. An upstream locus can condition what becomes available downstream without controlling every later transformation. This conditioning need not wait for either locus to produce Counterfactual Exclusion.

Counterfactual Inclusion Fields remain locus-indexed even where they depend upon shared physical conditions. The field of locus A is not merged with the field of locus B merely because their possibilities interact. Their branch relations may nevertheless overlap through shared resources, locations, interfaces, pathways, temporal windows, or other physical conditions. Some such relations may be co-realisation-incompatible: a branch may be genuinely included within A's Counterfactual Inclusion Field and another genuinely included within B's while the shared conditions do not permit both to be equivalently realised. Co-realisation incompatibility is a relation among potentials. It is not itself Counterfactual Exclusion, because no branch need yet have been realised.

One locus may therefore condition another's Counterfactual Inclusion Field while its own relevant power remains unapplied. If the possibilities available to B differ materially under the

standing organisation or condition supplied by A compared with an adequate condition in which A's relevant organisation is absent, disabled, withdrawn, relocated, narrowed, or otherwise changed, then A has conditioned B's possibility field. The difference may concern which branches remain included, how accessible they are, their probability, timing, persistence, cost, reachability, or the conditions under which later branching can occur.

Responsibility Provenance does not follow from overlap or incompatibility alone. It follows only where the field-conditioning relation is itself recoverable through distinction-conditioned organisation attributable to the evaluated locus. A passive obstacle may condition another locus's possibilities while remaining only part of causal provenance. A retained, maintained, positioned, exposed, withdrawn, or otherwise governed power-bearing organisation may carry Responsibility Provenance where the relevant discriminator is met.

This establishes only the pre-strategic causal groundwork. It does not yet require that either locus represent the other's field, anticipate the other's behaviour, optimise against it, or occupy a formal game space. Those stronger relations require later treatment.

This matters because responsibility relevance can diverge from physical proximity.

The final actor may be close to the consequence but operate inside a branch-space heavily conditioned by an earlier abstraction. An upstream locus may be physically distant yet have produced the category that determined which options were presented. Neither fact decides the attribution alone. The question is what distinction-conditioned transformations materially changed the branch.

A reliable responsibility account therefore resists the search for a single responsible noun. It reconstructs a path proportionate to the closure demand.

Where the path is simple, the account may be simple.

Where the path is deep, compressed, or distributed, the account must preserve enough provenance to prevent convenience from becoming false causation.

---

# 6. Nested loci, distributed provenance, and non-transfer

Responsibility becomes structurally difficult when one locus's output becomes another locus's input.

A finite locus does not receive reality as total state. It receives relation through an access path. The output of an upstream locus may therefore arrive already transformed. A sensor has classed a physical range. A person has summarised an observation. A model has generated a score. A record has retained a prior decision. A rule has compressed a category. A parent, teacher, clinician, engineer, administrator, or machine may then act upon that already-structured relation.

The receiving locus does not inherit the upstream frame as its own. It encounters a new relation under its own boundary, scale, horizon, discrimination regime, and closure demand.

If it performs a fresh distinction-conditioned transformation, a new responsibility event is created.

This gives the first law of distributed responsibility:

Provenance composes; responsibility does not simply transfer.

Suppose locus A discriminates a condition and produces output X. Locus B receives X and merely transmits it without a material distinction-conditioned transformation. The responsibility provenance of the relevant operation may remain concentrated upstream, although B may still contribute causal conditions such as transmission integrity.

Now suppose B classifies X, combines it with retained state, excludes an alternative, applies a threshold, or routes a later branch. B has introduced a new transformation. The resulting consequence may depend both on A's earlier distinction and on B's local use of it.

The correct output is not that responsibility moved from A to B.

Nor is it that both are equally responsible.

The correct output is that responsibility provenance now contains at least two localisable transformations whose contributions must be distinguished.

This yields three reciprocal guardrails.

First, local transformation does not erase upstream provenance. A downstream user cannot make an upstream classification cease to matter merely by acting upon it.

Second, upstream provenance does not absorb downstream transformation. The producer of an earlier distinction does not automatically become responsible for every later use, interpretation, recombination, or action performed by another locus.

Third, receipt does not equal inheritance. A locus that receives a relation does not automatically inherit the responsibility status of the locus that produced it. New attribution depends on what the receiver actually does with the relation within its own frame.

These rules apply below human language. A cellular pathway may receive the output of another regulatory pathway. A neural subsystem may receive a processed signal. A control system may act on another subsystem's state estimate. The nested structure is not first institutional. Institutions later make it deeper, more portable, and harder to recover.

Distributed provenance also means that one consequence can contain multiple responsibility-bearing paths without requiring a single total allocation. This is important. Ordinary responsibility discourse often asks how much responsibility belongs to each party, as though a conserved quantity must be divided among them. The primitive architecture does not establish such conservation.

One locus may be responsible for producing a false classification. Another may be responsible for relying on it despite available contradiction. Another may be responsible for designing a process in which no correction channel exists. Another may be responsible for a final application that transformed the inherited relation again. These are not necessarily slices of one scalar quantity. They are distinct responsibility relations connected to one consequential branch.

The same point protects against dilution. If many loci contribute, it does not follow that each becomes only weakly responsible by arithmetic division. A transformation can remain strongly attributable even inside a crowded causal network. Conversely, the mere presence of many actors does not establish that all have responsibility provenance.

The appropriate structure is therefore a graph of local transformations rather than a balance sheet of blame.[3]

This becomes especially important where abstraction accumulates. An early classification can narrow later branch-space. A retained record can make an earlier closure persist after its original context disappears. A formal rule can repeat the same transformation across thousands

of cases. A machine can scale a distinction-conditioned path far beyond the physical reach of the locus that first produced it. In such systems, causal proximity becomes an increasingly poor proxy for responsibility provenance.

But distance alone does not strengthen upstream responsibility either. The path must still show that the retained or transmitted abstraction remained materially operative in the later branch. If later loci corrected, replaced, ignored, or transformed it beyond recognition, the earlier contribution may weaken or change.

Nested responsibility therefore requires localisation at every material transformation.

No locus can be understood merely as the endpoint of what came before or the source of everything that came after.

---

# 7. Responsibility-relevant capacity and the Responsibility Gradient

Primitive responsibility establishes a low threshold. It does not imply that all responsibility-bearing loci are equivalent.

A thermostat can discriminate a narrow condition and route a narrow branch. A bacterium can integrate multiple environmental and internal distinctions over time. A nervous system can retain and combine past state, compare alternatives, and redirect behaviour. A human locus can often form explicit abstractions, represent uncertainty, seek additional evidence, anticipate downstream effects, reopen a prior closure, and describe reasons for action. A formal institution may distribute such operations across many people, records, rules, and machines.

The difference should not be described as generic complexity.

The relevant variable is responsibility-relevant capacity: the range and depth of operations through which a locus can discriminate, retain, abstract, compare, project, correct, preserve residual, govern its power-bearing conditions, and alter consequential branch relations.

Several dimensions are especially important.

Discrimination breadth concerns how many materially relevant alternatives can become operative for the locus. A binary threshold and a high-dimensional diagnostic system do not face the same branch field.

Discrimination resolution concerns how finely the locus can separate alternatives. A system that distinguishes only gross states may be unable to respond to a relation that a higher-resolution system can register.

Retention concerns whether prior distinctions remain available across time. A transient switch and a locus capable of retaining long causal history face different responsibility conditions because past relation may remain available to later correction.

Abstraction depth concerns how many transformations stand between target relation and operative class. A locus acting on raw contrast, a locus acting on a retained category, and a locus acting on a rule derived from earlier categories occupy different positions in the causal architecture.

Alternative representation concerns whether the locus can retain multiple represented alternatives rather than collapsing immediately to one. This matters because responsibility

changes where alternatives could have remained available or recoverable but were prematurely removed from consideration.

Residual capacity concerns whether unresolved relation can be preserved rather than silently converted into certainty. A primitive switch cannot be required to represent uncertainty it has no architecture to carry. A scientific, human, or institutional locus may have much stronger capacity to retain residual and state limits.

Comparator and correction capacity concern whether the locus can test its own transformation against mismatch, anomaly, alternative model, counterexample, or changed condition.

Projection capacity concerns whether material downstream consequences can be represented within the operative horizon. No locus is responsible for modelling the entire causal future. But where a known channel, repeated pattern, or explicit projection is available, it becomes part of the responsibility frame.

Control capacity concerns whether the locus can stop, delay, revise, reroute, escalate, request further input, or otherwise alter the branch after relevant uncertainty or mismatch becomes available.

Power-governance capacity concerns whether the locus can discriminate and alter the conditions through which its own Counterfactual Inclusion Field affects itself or other loci. Relevant operations may include maintaining, exposing, concealing, narrowing, expanding, disabling, delegating, relocating, withdrawing, monitoring, or otherwise changing a power-bearing capability or pathway. No locus is required to represent its Counterfactual Inclusion Field as a formal field. The question is whether the material parameters through which that potential affects later possibilities are themselves discriminable and governable within the locus's architecture.

Causal leverage concerns how strongly the local transformation conditions downstream branch-realisation. A tiny classification used once and the same classification embedded into a high-volume decision system do not have the same causal reach, even if the initial discriminative act is similar.

The Counterfactual Inclusion Field therefore gives the Responsibility Gradient an outer causal envelope without turning field size into a responsibility score. A larger CIF does not by itself establish greater responsibility. What expands the field of possible attribution is the recoverable capacity of the locus to discriminate and govern the parameters through which its power potential becomes consequential. As those capacities increase, a greater portion of the locus's power-in-potential may become responsibility-relevant even before any particular branch is exercised.

The gradient remains ordinal and relational. Capacity establishes what may become attributable; provenance establishes what is attributable. The relevant proposition is therefore not that responsibility is numerically proportional to the size of an IPL's Counterfactual Inclusion Field, but that responsibility may deepen and broaden as the IPL becomes capable of discriminating and governing a greater proportion of the consequential relations carried by that field.

These dimensions produce the Responsibility Gradient.

The Responsibility Gradient is not a moral ranking of beings. It is not a scalar blame score. It does not say that a more intelligent or sophisticated system is globally more responsible. It describes an ordinal and relational increase in the depth and breadth of responsibility attribution that can become available as distinction-conditioned causal capacity increases.[4]

The word can is essential.

Greater capacity expands the field within which responsibility may be attributed. It does not prove responsibility for a particular outcome. The actual claim still requires provenance.

A highly capable locus may have no responsibility provenance for an event it did not encounter or transform. A simple locus may carry strong primitive responsibility for a narrow branch that depends directly on its discriminator. A sophisticated institution may be less responsible for one local outcome than a small subsystem whose decision rule actually conditioned the branch.

The gradient therefore combines capacity with causal relation rather than replacing causal relation with capacity.

This is where responsibility begins to diverge from immediate signal path. As retained abstraction deepens, the mechanism that defines, preserves, or excludes distinctions can become more consequential than the physical carrier that transmits them. A later act may be physically proximal but heavily preconditioned. An earlier abstraction mechanism may be causally distant but repeatedly operative.

Again, this does not license automatic upstream blame. It changes the question. Instead of asking who touched the outcome last, responsibility analysis asks which loci had which responsibility-relevant capacities and which of their distinction-conditioned transformations actually shaped the branch.

The gradient is therefore not an escape from provenance.

It is the architecture that explains why provenance becomes richer as finite systems become capable of carrying more of the relations they transform.

---

# 8. The immediate-frame principle: access, residual, and available alternatives

Responsibility must be evaluated from within the conditions under which the locus could actually operate.

This follows from the finite-observer architecture. No locus possesses total state. A system acts through relations available across its boundary, interface, scale, horizon, discrimination regime, abstraction, registration, retention, and redeployment conditions. An exterior observer may later reconstruct a wider field. That reconstruction can improve causal understanding. It cannot be retroactively inserted into the earlier locus as though the locus possessed it at the time.

This gives the Immediate-Frame Principle:

Responsibility attribution is bounded by the relations, distinctions, alternatives, residual, projection channels, and corrective operations available or recoverable to the evaluated locus within its own indexed frame at the relevant time.

The principle is not an excuse doctrine. It is an anti-anthropomorphism and anti-hindsight discipline.

A thermostat cannot be faulted for failing to preserve explanatory residual if its architecture contains no operation by which residual could be represented. A bacterium cannot be held to an exterior evaluator's represented alternative set merely because a scientist can describe

additional strategies after observation. A person cannot be assigned responsibility for a hidden causal mechanism solely because later science discovers it. A machine cannot be treated as though it possessed the world-model of its external evaluator if that relation never entered its operative architecture.

But the same principle works in the opposite direction.

Where a relevant distinction was available, where a contradiction was registered, where an uncertainty marker was retained, where a stop condition could be triggered, where an alternative remained available or recoverable within the locus's operative frame, or where a correction pathway was available, those relations are part of the frame. A locus cannot invoke finitude to erase operations it actually possessed.

This is especially important for residual.

Residual is not a demand that every system represent everything it does not know. That would be impossible. Residual becomes responsibility-relevant where unresolved relation has entered a form the locus can retain, mark, bracket, monitor, or use to constrain closure.

A primitive locus may have little capacity for explicit residual. Its responsibility is correspondingly narrow. A more capable locus may encode confidence, uncertainty, failed checks, contradictory evidence, unresolved alternatives, incomplete records, exception states, or known model limits. Once such residual is available, the responsibility relation changes. Closing as though the residual were absent is a different transformation from closing where no residual was representable.

The same is true of alternatives. Responsibility does not require awareness of the whole Counterfactual Inclusion Field. It requires attention to the alternatives and field-conditioning relations available or recoverable within the locus's Immediate Frame. A locus may face only two operative alternatives. Another may face many. A third may be able to suspend, defer, ask, escalate, narrow, withdraw, or alter a standing capability rather than choose among the apparent primary branches. An alternative may be included within the Counterfactual Inclusion Field without becoming responsibility-relevant merely because it exists there.

Non-application must be handled by the same rule. The mere inclusion of an unused branch within the Counterfactual Inclusion Field does not make a locus responsible for failing to realise it. Non-application becomes responsibility-bearing where the relevant alternative was available or recoverable within the locus's Immediate Frame, its organisation could apply, withhold, stop, redirect, or otherwise govern the relevant distinction, and the non-application materially conditioned the consequential branch relation relative to an adequate comparator under the relevant temporal conditions.[5]

Standing field-conditioning is distinct from non-application. A locus may never exercise a particular capability and may never produce Counterfactual Exclusion through it, while the continued existence, maintenance, exposure, location, or other governed condition of that capability materially changes another locus's Counterfactual Inclusion Field. Where that field-conditioning effect was itself available or recoverable within the first locus's Immediate Frame, and where its organisation could maintain, alter, withdraw, narrow, or otherwise govern the relevant condition, Responsibility Provenance may attach to the standing relation. The attributed consequence is then the shaping of another locus's possibility field, not the failure to realise the unused branch.

Projection is similarly bounded. No system can be responsible for consequences requiring arbitrary prediction depth. The relevant question is whether a material downstream relation was available through established channel, repeated pattern, explicit model, prior warning, or other recoverable path under the horizon being claimed.

The Immediate-Frame Principle therefore prevents two opposite failures.

The first is exterior over-attribution: an evaluator treats its own larger abstraction as though it had been available to the target and then assigns responsibility for failure to use it.

The second is interior laundering: a capable locus narrows its described frame after the fact, ignoring distinctions, residual, alternatives, warnings, or correction paths that were in fact available to it.

Responsibility audit must resist both.

The frame is indexed to the locus, but it is not whatever the locus later says it was. It is a recoverable technical claim about availability, capacity, and operation under the relevant conditions.

This keeps responsibility realist and finite at the same time.

The locus is not judged against omniscience.

It is judged against the relations its own organisation could actually make consequential.

---

# 9. Responsibility through time: retention, hindsight, and update

Responsibility is temporal because distinction-conditioned organisation acts through time.

A present consequence may depend on a distinction registered moments earlier, a category retained for years, a rule inherited across generations, or a model trained on data from another period. The relevant responsibility provenance may therefore be distributed not only across loci but across horizons.

Retention matters because it allows an earlier abstraction to remain causally operative after the conditions of its formation have changed. A retained state can become a standing input to later branch-realisation. A record can carry a closure forward. A category can persist after its target relation shifts. A learned threshold can continue to route cases after the environment changes.

The existence of retention does not make the earlier locus permanently responsible for every later consequence. The provenance question remains local. Did the retained distinction materially condition the later branch? Did later loci merely transmit it, or did they re-evaluate, transform, replace, or correct it? What update channels existed? Which loci controlled them?

Temporal responsibility therefore requires a distinction between origin, persistence, redeployment, and update.

The origin locus may be responsible for forming the abstraction under its own immediate frame.

A retention locus may be responsible for preserving it under conditions that made persistence consequential.

A redeployment locus may be responsible for applying it to a new case.

An update locus may be responsible for revising, suspending, or retaining it after materially changed relation becomes available.

These functions may belong to one system or many.

Persistence may also be consequential before later redeployment. A retained or maintained power-bearing organisation can remain a standing condition of another locus's Counterfactual Inclusion Field even where the relevant capability is never exercised. The provenance question is then not whether the first locus produced Counterfactual Exclusion, but whether its distinction-conditioned organisation materially altered the second locus's branch availability through persistence, and which locus could discriminate, maintain, revise, relocate, narrow, withdraw, or otherwise govern that condition over the relevant horizon.

This permits responsibility to persist without turning persistence into permanent ownership. An origin locus does not remain responsible for every future field effect merely because it once created the condition. A persistence or update locus may introduce new Responsibility Provenance where it actively or knowingly maintains the relation under changed conditions. Conversely, where no relevant discriminator or corrective capacity was available, continued physical presence may remain causal background rather than responsibility-bearing persistence.

This structure blocks hindsight error.

Later falsification is not sufficient proof of earlier responsibility failure.

A model can be wrong without having been irresponsibly formed. A decision can produce harm without the relevant causal relation having been available at the time. A provisional closure can later fail under new evidence even though it was the strongest available closure under earlier conditions.

Retrospective responsibility requires more than saying that the outcome turned out badly or that later knowledge would have supported another branch.

It requires localisation of an earlier available or recoverable failure condition.

Was contradictory relation already available? Was material residual suppressed? Was a known comparator ignored? Was an update trigger crossed? Was a relevant distinction registered but not retained? Was an alternative intentionally or structurally removed from consideration despite remaining available or recoverable within the locus's operative frame? Was a correction channel available but blocked? Did the locus continue to redeploy an abstraction after its reference conditions materially shifted?

Where the answer is no, later evidence changes the present model and triggers update. It does not manufacture earlier responsibility.

Where the answer is yes, the retrospective claim may strengthen because the failure can be located inside the earlier locus's own operative frame.

This distinction is especially important in recursive systems. Once records, rules, classifications, or models are retained, the responsibility architecture includes not only the original transformation but the governance of persistence. A system capable of retaining a distinction may also require a way to recognise when the conditions that grounded retention have weakened.

The burden is proportional. A simple locus with no endogenous update capacity cannot be assigned an update duty it cannot perform. A sophisticated system that explicitly monitors drift, exceptions, contradiction, or model failure occupies a different responsibility position because those relations are part of its architecture.

Responsibility through time is therefore not permanent ownership of consequences.

It is provenance across formation, retention, redeployment, and correction.

The past matters only insofar as its distinctions remain causally operative in the present branch.

---

# 10. Responsibility Warrant and Responsibility Distortion

Primitive responsibility does not exhaust ordinary responsibility.

Human systems must coordinate action across causal architectures too complex, distributed, and incomplete to reconstruct from first principles every time. They therefore allocate duties, answerability, supervision, liability, stewardship, decision rights, and burdens. Parents are assigned responsibilities toward children. Professionals are assigned responsibilities within domains. Employers are assigned responsibilities for workplace conditions. Directors are assigned responsibilities for governance. Public bodies are assigned responsibilities for services. Operators are assigned responsibilities for systems they did not design. Institutions create named accountable roles so action does not disappear into distributed causation.

These allocations are not primitive responsibility provenance.

They are Responsibility Warrants.

Responsibility Warrant is an authorised or normatively operative allocation of answerability, obligation, burden, liability, stewardship, supervision, or decision ownership to a locus under a specified social, moral, legal, organisational, procedural, or institutional regime.

The word warrant is used here in the attributional sense. A regime supplies a reason why a locus is to carry some responsibility burden for a purpose. That reason may be contractual, legal, moral, organisational, protective, practical, or administrative. The warrant may deliberately assign responsibility beyond the locus's direct causal provenance. A duty of care can require action before harm occurs. Strict liability can allocate loss without proving fault. Supervisory responsibility can attach to a role because the system needs a locus with authority to intervene. Parental responsibility can allocate duties of provision, attendance, protection, or support without claiming that the parent causally produces the child's total educational state.

Such warrants may be necessary.

The distortion begins when the warrant is allowed to travel as though it established the underlying causal provenance.

Responsibility Warrant does not constitute Responsibility Provenance.[6]

A society may declare that parents are responsible for a child's education. As a warrant, the claim can assign concrete duties: provide access, support attendance, respond to school communication, or make certain decisions. As a provenance claim, however, the same sentence is radically under-specified. A child's educational state is produced through the child's own retained distinctions, teachers, peers, curriculum, school structures, assessment systems,

resources, family environment, language, technology, prior history, and many other relations. The parent may carry strong provenance for some branches and little for others. A social warrant that allocates broad answerability does not turn that distributed causal architecture into a single-parent provenance relation.

Responsibility Distortion occurs where an abstracted responsibility attribution or warrant fails to preserve the material responsibility provenance relevant to the claim or action for which the attribution is used, while continuing to carry stronger responsibility language as though it did.

Several recurrent forms follow.

Proximity distortion assigns responsibility to the actor nearest the visible outcome because the causal path immediately before the consequence is easiest to see.[7]

Role distortion treats formal office, title, family relation, professional identity, or organisational position as if the role itself established the causal transformations that produced the outcome.

Container distortion attributes a distributed process to the most familiar containing entity: the parent, company, hospital, school, algorithm, government, or individual operator.

Outcome distortion infers responsibility from success or failure after the event rather than from the distinction-conditioned path available at the time.

Authority distortion treats legal, institutional, procedural, or social assignment as proof of causal provenance.

Provenance compression reduces a multi-locus transformation path into one accountable noun for practical use and then forgets that the noun was a compression.

Warrant laundering occurs when one kind of warrant is spent as another. A locus may have operational responsibility to act, yet that is narrated as causal authorship. A role may have custodial responsibility for a record, yet later users treat the role as the source of the record's epistemic strength. A legal liability rule may allocate loss, yet later discourse treats liability as proof of causal fault.

None of these failures implies that warrants should disappear. Distributed systems require Soft Horizons for action. Someone may need authority to decide, a duty to inspect, an obligation to protect, or liability to ensure repair even where provenance is incomplete.

Where a Responsibility Warrant authorises, requires, or allocates consequential action while provenance remains incomplete, the warrant must not erase the action-under-residual structure. The action threshold, affected locus, material residual, safeguard, monitoring condition, stop condition where applicable, and update trigger should remain recoverable. Necessity of allocation may justify action; it does not convert unresolved provenance into causal closure.

The discipline is narrower and stronger.

The warrant must remain identifiable as warrant.

Its scope, purpose, authority, and residual relation to provenance must not disappear.

A legitimate warrant can differ from provenance without pretending to be provenance.

A distorted warrant hides that difference.

# 11. Cross-scale demonstrations

The primitive definition must survive movement across scale without flattening the cases.

The passive causal case comes first.

A stone rests against a door and prevents it from closing. The stone is causally relevant. Remove it and the branch changes. But the relevant difference is not mediated by an organised discriminator belonging to the stone. The stone does not register open versus closed and route later state through that distinction. Under the technical definition used here, causal provenance is present while primitive responsibility provenance is not established.

A thermostat provides the minimal positive case.

Temperature relative to a threshold becomes an operative alternative. The sensing-and-switching organisation routes later behaviour differently depending on the registered condition. Disable the relevant pathway and the output ceases to track the distinction in the same way. The thermostat therefore carries primitive responsibility provenance for the narrow switching transformation under the stated frame.

The claim must remain narrow. The thermostat is not thereby responsible for the design of the building, the choice of set point, energy policy, occupant comfort, system maintenance, or downstream social consequences. Those relations belong to other loci and other transformations. Primitive responsibility does not inflate a simple discriminator into general agency.

A living cell supplies a richer case.

A cell may discriminate molecular shape, ion concentration, nutrient availability, damage state, membrane condition, stress signal, or regulatory threshold. These distinctions can be retained across biochemical state, integrated with other pathways, and redeployed into repair, transport, movement, expression, metabolism, or apoptosis. Responsibility provenance becomes correspondingly richer because many local distinctions condition many branches inside a maintained organisation.

Still, no moral agency has been established. The cell is not blamed for a failed repair. Its responsibility is causal and indexed. If a repair pathway discriminates damage and routes a checkpoint, the relevant transformation can be attributed to that organised locus. If the pathway lacks access to a particular defect, the Immediate-Frame Principle blocks exterior over-attribution.

A human decision locus adds abstraction, memory, explicit alternatives, residual representation, and correction capacity.

Consider a clinician deciding whether to treat. The present branch may depend on symptoms, measurements, prior records, diagnostic categories, known side effects, uncertainty, patient condition, and available alternatives. The clinician may be able to seek further evidence, delay, consult, disclose uncertainty, or act under urgency. Responsibility provenance therefore includes not merely the final prescription but the distinctions available, the abstractions applied, the residual retained or suppressed, and the branches the clinician could actually select.

A later adverse outcome does not prove earlier responsibility failure. If the causal mechanism was unavailable and the decision preserved known residual under a proportionate threshold, the later discovery triggers update. If warning signs were available, ignored, or actively excluded despite a correction path, the provenance claim changes.

A distributed educational case shows the difference between provenance and warrant.

A child's educational state develops through many loci. The child registers and retains distinctions. Teachers select explanations, feedback, and assessment. Curriculum designers structure available categories. Schools allocate time, resources, and interventions. Peers alter social and informational conditions. Families shape access, language, expectations, routines, and support. Platforms and records may mediate further distinctions. No one of these descriptions proves how much any locus contributed in a particular case, but the architecture is plainly distributed.

A social regime may nevertheless state that a parent is responsible for the child's education. That statement can function coherently as a warrant assigning specified obligations. It becomes distorted when it is used as if the warrant demonstrated that the parent causally produced the child's educational attainment or failure across the distributed path.

The distinction matters in both directions. Provenance analysis may reveal substantial parental contribution to some branches. It may also reveal causal transformations located primarily elsewhere. The point is not to exonerate or accuse the parent. It is to stop an administrative abstraction from substituting for causal mapping.

A machine-mediated case provides the boundary to the next Part.

Suppose a model classifies an applicant, an operator receives the classification, an institutional rule determines how the score is used, and a final decision affects the applicant. Responsibility provenance may involve the training and feature architecture, the model's classification, the interface design, the operator's available alternatives, the institutional threshold, and the final application. The final click is not automatically the whole responsibility path. The model is not automatically the whole path either.

If the operator could challenge the score and contrary evidence was visible, local responsibility may arise at the reliance transformation. If the interface concealed the uncertainty and permitted no meaningful override, the branch-space was different. If the institution mandated automatic use, the rule itself becomes a major responsibility-bearing abstraction. If the model producer had no control over the later deployment and the receiving institution transformed the output into a new use outside the original frame, upstream provenance does not automatically absorb the downstream decision.

Across all of these cases the same discipline holds.

Find the operative distinction.

Find the organisation that made it consequential.

Find the comparator.

Find the locus.

Recover the path far enough to support the claim.

Do not let familiarity decide the attribution before the relation has been mapped.

---

# 12. From Responsibility to Trust

Responsibility is causally operative before trust.

A thermostat can carry primitive responsibility for threshold-conditioned switching without trusting anything. A cell can perform responsibility-bearing discrimination without first relying on another locus in the richer trust sense. A finite locus can transform available distinction and thereby acquire responsibility provenance before any question of vicarious closure arises.

But distributed responsibility creates the condition under which trust becomes necessary.

Once one locus's transformed output becomes another locus's available relation, the receiver faces a practical problem. It cannot continuously reconstruct every upstream distinction, comparator, abstraction, residual, retention condition, and causal pathway before every use. As the number of loci grows and abstraction becomes deeper, total repeated provenance reconstruction becomes impossible for a finite system.

The receiver must therefore decide how much of the upstream relation to reopen.

Sometimes it will audit directly. It may repeat a measurement, inspect a record, test a model, verify a source, challenge a classification, or reconstruct a causal path.

Sometimes it will rely.

That reliance does not make upstream responsibility disappear. Nor does it transfer upstream responsibility to the receiver. It creates a new local transformation: the receiver permits an inherited relation or carrier to satisfy enough of its own closure demand that branch-realisation can continue.

Trust begins there.

This shows why responsibility must precede trust in the primitive sequence.

Responsibility tells us that inherited relations are not neutral surfaces. They may already contain transformations performed by other loci. Their present form may be the result of prior discrimination, retention, abstraction, exclusion, correction, or misclosure. When a receiving locus relies on them, it acts across an already responsibility-bearing path.

Trust then becomes an economy over responsibility provenance.

It reduces the work of repeated audit by allowing some upstream closure-work to remain backgrounded.

This also clarifies the boundary between trust and excuse.

"I trusted the source," "I trusted the model," "I trusted the expert," or "I trusted the institution" may explain why a locus did not reconstruct the full upstream path. The statements do not by themselves determine responsibility for the downstream transformation.

The responsibility question remains local.

What relation was available to the relying locus? What residual was visible? What anomaly should have reopened audit? What alternatives existed? What reliance threshold was applied? What transformation did the locus perform after accepting the carrier? What control did it have over the next branch?

Trust can therefore reduce audit without exporting responsibility.

An upstream source remains responsible for its own responsibility-bearing transformations. A downstream relying locus remains responsible for its local transformation insofar as the conditions of primitive or richer responsibility are met. The trust relation connects the paths without collapsing them.

This also explains why betrayal and repair later matter. Betrayal is not simply a bad outcome delivered by a trusted carrier. It concerns a relied-upon relation whose responsibility-bearing conditions fail in a way that matters to the relying locus. Repair requires more than renewed confidence. The path by which responsibility and reliance were misaligned must become recoverable enough for trust to be recalibrated.

The handoff is therefore precise.

Responsibility asks where consequential transformation is attributable.

Trust asks when a finite locus may allow another relation, carrier, or closure process to remain below immediate audit while continuing to act under residual.

Responsibility supplies the causal provenance that trust can background but cannot abolish.

Trust supplies the branch-space economy required once responsibility-bearing relations become too distributed to reconstruct continuously.

The next paper can therefore begin from a stronger primitive problem than trust alone.

A finite locus must rely across paths whose transformations were performed elsewhere and whose full provenance cannot be reopened at every use.

Trust is the finite mechanism that makes such continuation possible.

---

# 13. The discipline of Responsibility

A primitive responsibility theory is useful only if it resists the familiar shortcuts by which attribution becomes moralised, anthropomorphised, institutionalised, or detached from causal relation.

The first discipline is to preserve the causation boundary.

Not every cause is responsibility-bearing. Passive constraint, background condition, magnitude, complexity, correlation, and external description do not replace the NTC-derived discriminator. Where no operative distinction-conditioned organisation can be localised, the correct output may remain causal contribution without primitive responsibility.

The second discipline is to preserve the locus.

Responsibility must not float across convenient nouns. The person, organisation, machine, family, school, government, platform, or environment may be too wide, too narrow, or incorrectly framed. The locus is the organisation whose transformation is being attributed.

The third discipline is to preserve the distinction.

A responsibility claim must identify what alternative mattered to the locus. If the only distinction exists in the evaluator's description, the attribution is not yet earned.

The fourth discipline is to preserve the comparator.

Responsibility claims should weaken where the downstream branch would remain materially unchanged without the proposed distinction-conditioned organisation. Proximity, confidence, role, and outcome do not substitute for comparator pressure.

The fifth discipline is to preserve provenance.

A final action must not erase the abstractions that shaped it. An upstream abstraction must not absorb every downstream transformation. Distribution must not become either dilution or totalisation. Provenance composes through local transformations.

The sixth discipline is to preserve the Immediate Frame.

No locus should be assigned responsibility for relation that was unavailable to its architecture merely because an exterior observer can reconstruct it later. Equally, no capable locus should hide available residual, alternatives, warnings, correction routes, or update signals by retrospectively narrowing its frame.

The seventh discipline is to preserve temporal conditions.

Later falsification is not proof of earlier responsibility failure. Retrospective attribution requires an earlier available or recoverable failure condition. Update changes the present responsibility landscape; it does not rewrite the past without evidence.

The eighth discipline is to preserve the gradient without turning it into a score.

Responsibility-relevant capacity can become richer across discrimination, retention, abstraction, residual handling, projection, correction, and control. Those dimensions expand the possible field of attribution. They do not create a universal quantity of responsibility that can be compared across all systems.

The ninth discipline is to preserve the provenance-warrant boundary.

Roles, laws, duties, policies, family relations, professional standards, and institutional procedures may create Responsibility Warrants. Such warrants can be legitimate even where they deliberately simplify or depart from causal provenance for a specified purpose. They become distorted when they are narrated as though the authority that assigned them also established the causal path.

The tenth discipline is non-export.

A locus cannot discharge responsibility for its own consequential distinction-conditioned transformation merely by invoking an upstream source, downstream receiver, authority, procedure, carrier, model, expected interpretation, or future use.

The reciprocal rule is equally important. A downstream transformation does not erase upstream provenance, and receiving an upstream relation does not automatically transfer the upstream locus's responsibility to the receiver.

These disciplines do not make responsibility easy.

They make it localisable.

A finite responsibility account will still carry residual. The causal path may be incomplete. The correct locus may be contested. Comparator construction may be difficult. Multiple abstractions may interact. Some relevant alternatives may be unrecoverable. Institutional warrants may be necessary before provenance is settled. Action may proceed before the responsibility graph is complete.

The method required is therefore not total attribution before action.

It is disciplined attribution without hidden strengthening.

Where responsibility is primitive, say what distinction-conditioned branch is attributable.

Where responsibility is distributed, recover the material transformation path.

Where capacity is limited, state the frame.

Where residual remains, retain it.

Where a warrant assigns duty beyond provenance, identify the warrant as warrant.

Where later evidence changes the path, update without laundering hindsight into earlier failure.

The governing rule can now be stated in its complete form:

Responsibility must not carry attribution-strength beyond the recoverable distinction-conditioned causal contribution of the locus under the reference, comparator, capacity, and temporal conditions being claimed. Where responsibility is assigned by warrant, the warrant must not be treated as provenance merely because authority, role, convention, procedure, or institution supplies it.

This rule does not abolish moral responsibility, legal responsibility, parental responsibility, professional responsibility, or institutional accountability.

It gives those later abstractions a causal ground against which their claims can be examined.

---

# 14. Conclusion: Responsibility without blame

Responsibility begins earlier than its ordinary language suggests.

It does not begin when a society assigns duty. It does not begin when a court assigns liability. It does not begin when a person forms an intention, when an institution names an accountable role, or when an observer praises or blames an outcome.

Those are later responsibility regimes.

The primitive begins where distinction-conditioned organisation makes a consequential branch relation depend on an operative or retained alternative and where that dependence can be attributed to a locus under specified reference and comparator conditions.

Causation is the substrate. Non-Trivial Computation supplies the discriminator. Power identifies the Counterfactual Inclusion Field: the indexed branch-changing potential made physically available by distinction-conditioned organisation. Counterfactual Exclusion identifies the branch-realisation relation produced where some part of that potential is applied. Responsibility is the attributional consequence, but it is not restricted to Counterfactual Exclusion. It asks which realised branch, non-application, retained condition, maintenance relation, or cross-locus field-conditioning effect of that potential belongs to the evaluated locus under the actual conditions.

This makes primitive responsibility both thinner and more general than the familiar human noun. A thermostat can carry primitive responsibility for a narrow switching transformation without agency in the moral sense. A cell can carry richer responsibility provenance across multiple distinction-conditioned pathways without blame. A human locus can carry further responsibility because more alternatives, residual, abstraction, projection, correction, and control may be available. A distributed system can carry responsibility across many loci without requiring the causal path to collapse into one responsible party.

The central object is Responsibility Provenance.

Responsibility provenance recovers the path by which local transformations make a downstream branch relation depend on distinction-conditioned organisation. It separates

responsibility from passive causation, external description, causal proximity, role, and outcome. It makes clear that the final physical act may not contain the whole responsibility architecture and that an upstream abstraction may remain causally operative without owning everything that follows.

Provenance composes.

Responsibility does not simply transfer.

A new local transformation can create new responsibility provenance. It does not erase what came before. An earlier contribution can remain material. It does not absorb every later use. Distribution therefore produces a graph of attributable transformations rather than a scalar quantity to be divided among named actors.

The Responsibility Gradient explains why this graph changes with capacity. A locus able to discriminate more alternatives, retain them, abstract them, preserve residual, compare branches, anticipate material consequences, correct, stop, update, and govern the conditions through which its own power affects other loci occupies a richer responsibility field. A wider Counterfactual Inclusion Field alone does not establish greater responsibility. Greater recoverable capacity to discriminate and govern consequential relations within and across that field expands what may become attributable. Provenance remains controlling.

The Immediate-Frame Principle prevents the gradient from becoming retrospective omniscience. A locus is evaluated against relation available or recoverable within its own indexed frame at the relevant time. Later discovery triggers update. It establishes earlier responsibility only where an earlier available failure condition can be localised.

This also establishes a boundary for later treatment of interacting power-bearing loci. Counterfactual Inclusion Fields may overlap through shared physical conditions, and branch relations within separately indexed fields may be co-realisation-incompatible. One locus may condition another's field without either yet producing Counterfactual Exclusion. Responsibility can already recover the attributable standing, retained, or governed relations that shape those possibilities. A later theory may add strategic representation, anticipation, competition, cooperation, or formal game spaces without having to reconstruct their causal and responsibility ground from scratch.

Responsibility Warrant then explains how the familiar social concept emerges.

Finite societies and institutions cannot reconstruct total causal provenance before every decision. They allocate duties, answerability, burdens, liabilities, stewardship, and decision ownership. Such warrants may be necessary, protective, practical, or just. They are not made false merely because they are abstractions.

But warrant is not provenance.

A parent can carry a social duty toward a child's education without thereby becoming the causal producer of the child's educational state. A professional role can carry an obligation to act without proving that the role caused the condition. An institution can allocate liability without discovering the complete responsibility graph. A named accountable locus can be useful for governance while remaining only a partial model of distributed causal provenance.

Responsibility distortion begins when that difference disappears.

The warrant becomes the cause. The role becomes the provenance. The final actor becomes the whole path. The institution's allocation becomes evidence that the allocated locus generated the outcome. At that point a necessary abstraction has hardened into causal overclaim.

The discipline of responsibility is therefore not a demand that all responsibility warrants mirror causal provenance exactly.

It is a demand that they not pretend to do so where they do not.

The causal sequence now reaches its next problem.

Once responsibility-bearing transformations are nested across loci, no finite receiver can reconstruct every upstream distinction, abstraction, comparator, residual, and update condition before every further branch. A system that attempted total provenance audit before each act would become non-functional.

Some inherited relations must therefore be allowed to remain backgrounded.

The receiving locus must sometimes rely.

That reliance does not create responsibility. It occurs across a responsibility-bearing causal architecture already in place.

This is why responsibility precedes trust.

Responsibility identifies where consequential transformation is attributable.

Trust explains how finite systems reduce repeated audit across such attributable dependency while residual remains.

The origin of responsibility is therefore not the origin of blame.

It is the point at which a difference made consequential by organisation becomes attributable to the locus that made it matter.

# Endnotes

[1] Moral-responsibility boundary comparator: Carolina Sartorio, “How to Be Responsible for Something Without Causing It” (2004), argues that moral responsibility need not entail causing the outcome for which an agent is responsible. This provides useful adversarial pressure against reading the primitive causal claim here as a reduction of every mature responsibility relation to causation. The present paper makes the narrower claim that primitive Responsibility is causal-attributional; later moral, legal, social, and institutional Responsibility Warrants may allocate responsibility on additional grounds and are explicitly separated from Responsibility Provenance.

[2] Responsibility-provenance adjacency and independent convergence: Jinwei Hu, Xinmiao Huang, Qisong He, Youcheng Sun, Yi Dong, and Xiaowei Huang, “Responsible Agentic AI Requires Explicit Provenance” (2026), argue that responsibility in agentic AI requires quantifiable and traceable provenance across the system lifecycle. The correspondence is functionally significant because both accounts make recoverable transformation paths a condition of disciplined responsibility attribution. Their target is agentic-AI governance and a computable causal-attribution architecture; Responsibility Provenance here is proposed as a more primitive locus-indexed relation across biological, human, machine, and institutional cases and does not inherit their responsibility tensor or agentic framework.

[3] Distributed-responsibility comparator: Dennis F. Thompson, “Moral Responsibility of Public Officials: The Problem of Many Hands” (1980), identifies the difficulty of assigning responsibility where many officials contribute to one outcome and criticises simple hierarchical or collective solutions that distort individual responsibility. This is a strong later-domain comparator for the present refusal to collapse a distributed path into one responsible noun or a conserved scalar quantity. Thompson works within moral and political responsibility and uses causal and volitional criteria; the present architecture is prior to that allocation problem and represents distributed Responsibility Provenance as a graph of local distinction-conditioned transformations.

[4] Graded-responsibility comparator: Hana Chockler and Joseph Y. Halpern, “Responsibility and Blame: A Structural-Model Approach” (2004), extend structural-model accounts of actual causation with a quantitative degree of responsibility and distinguish that degree from blame by adding an agent’s epistemic

state. The adjacency is important because both accounts reject undifferentiated causal attribution and permit responsibility relations to vary in strength. The Responsibility Gradient developed here is not their scalar degree of responsibility: it is an ordinal, relational architecture of responsibility-relevant capacity, while actual Responsibility still depends on recoverable provenance rather than on a universal numerical allocation.

[5] Omission and non-application comparator: Laura Christina Anne Fearnley, "Causal Responsibility Through Omission" (2026), asks what distinguishes omissions that make an agent causally responsible from mere failures to act and develops a robustness/invariance condition over the relevant causal chain. This is directly adjacent to the present refusal to infer responsibility merely because an unused alternative existed. Non-application here is responsibility-bearing only where the alternative was available or recoverable within the locus's Immediate Frame, the locus could govern application or withholding, and the non-application materially conditioned the consequential relation under comparator pressure; the account is therefore broader than omission by a mature agent and remains tied to the paper's primitive locus architecture.

[6] Responsibility-taxonomy comparator: H. L. A. Hart, Punishment and Responsibility (1968), distinguishes several senses of responsibility, including causal, role, liability, and capacity responsibility. That taxonomy is an important precursor to the need to prevent causal contribution, institutional role, answerability, and liability from collapsing into one undifferentiated noun. The present Provenance/Warrant distinction is not inherited from Hart: Responsibility Provenance is restricted to distinction-conditioned causal contribution under the paper's discriminator, while Responsibility Warrant covers authorised or normatively operative allocations that may legitimately extend beyond direct causal provenance.

[7] Responsibility-distortion comparator: Madeleine Clare Elish, "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction" (2019), describes complex automated systems in which moral and legal responsibility may be disproportionately assigned to the nearest human operator despite limited control over the wider system. This is a strong sociotechnical specialisation of the proximity, role, and container distortions identified here. The present account generalises the structural problem beyond human-robot systems: causal proximity and accountable role are not themselves Responsibility Provenance, and the relevant audit must recover the distinction-conditioned transformations distributed across the path.

# References

Chockler, H., & Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research*, 22, 93-115. https://doi.org/10.1613/jair.1391

Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society*, 5, 40-60. https://doi.org/10.17351/ests2019.260

Fearnley, L. C. A. (2026). Causal responsibility through omission. *American Philosophical Quarterly*, 63(3), 281-294. https://doi.org/10.5406/21521123.63.3.06

Hart, H. L. A. (1968). *Punishment and responsibility: Essays in the philosophy of law*. Clarendon Press.

Hu, J., Huang, X., He, Q., Sun, Y., Dong, Y., & Huang, X. (2026). Responsible agentic AI requires explicit provenance. *arXiv*:2605.17169. https://doi.org/10.48550/arXiv.2605.17169

Sartorio, C. (2004). How to be responsible for something without causing it. *Philosophical Perspectives*, 18, 315-336. https://doi.org/10.1111/j.1520-8583.2004.00031.x

Thompson, D. F. (1980). Moral responsibility of public officials: The problem of many hands. *American Political Science Review*, 74(4), 905-916. https://doi.org/10.2307/1954312