MINIMAL ETHICAL DIMENSIONS

Toward a Grounded, Plural Ethics

Working Paper — Shaky Foundation — 2026

Draft 0.2 — MIT Licence

Abstract

The dominant traditions in Western normative ethics share a structural assumption that is rarely examined as such: that the moral domain is ultimately organised around a single evaluative axis, and that the task of ethical theory is to identify which axis is fundamental. This document argues that the assumption is a category error, and that many non-Western traditions (Confucian, Buddhist, Ubuntu, Indigenous) were already structurally plural before the Western pluralist critique began. The world's ethical traditions are not competing answers to the same question. They are partial instruments, each measuring different dimensions of an underlying moral structure that is irreducibly plural.

The document proposes a diagnostic and constructive framework built on a central primitive: harm, understood as structural or functional degradation of something real. It rests on a single stated normative presupposition: that such degradation is pro tanto bad. It identifies eight candidate dimensions along which harm can occur (valence, agency, coherence, stability, relational integrity, coordination, information integrity, and conflict), subjects each to tests for observability, independence, and measurability, and maps the world's major ethical traditions onto those dimensions to show what each tracks, where it is strong, and where it is blind. It applies the same diagnostic standard destructively, eliminating concepts (brute moral wrongness, "the good" as an unexplained unity, purity) that do not meet its tests. The framework is stronger as a diagnostic for harm than as a complete description of positive goods, and it acknowledges this limitation honestly.

The framework argues that the convergence of historically independent traditions on the same dimensions is evidence that the dimensions are robustly useful across contexts and that they track features of entities and structures stable enough to be detected independently. It treats time not as a background condition but as a structural axis of the harm space, categorically different from the harm dimensions (which are axes along which harm occurs) but requiring equal analytical seriousness (as the axis along which harm profiles evolve). Every decision produces a harm profile across multiple dimensions, across time, under uncertainty, and no case is ever closed: every trajectory remains open on every dimension.

A five-step structured method for moral reasoning is developed: construct the harm profile, eliminate dominated options, check against non-negotiable floors, check the temporal trajectory, and make the residual tradeoff explicit. The method produces structured narrowing rather than terminal verdicts, resolving most decisions through its structural steps and narrowing the genuinely contested cases to a smaller set than they initially appear to be. Plurality is shown to constrain rather than dissolve: intra-dimensional ordering, floors, dominance, temporal analysis, and the diagnostic standard together produce substantive moral judgment without requiring single-axis reduction.

The method is applied to nine cases, each tracking all eight dimensions at multiple timesteps: civilisational catastrophe and its ongoing institutional legacy (Hitler), genuine multi-dimensional tradeoff with partial repair over time (Henrietta Lacks), information-correction with temporal inversion (Pentagon Papers), ecological collapse with divergent recovery trajectories (Aral Sea), contested environmental trajectory with the tipping point ahead (Amazon), coordinated multi-dimensional harm-reduction with compounding benefits (smallpox eradication), industrial-scale harm to non-human entities sustained by exclusion from the accounting (animal agriculture), simultaneous harm and benefit through the same mechanism at digital scale (social media), and unprecedented harm-reduction coexisting with civilisational-scale risk and contested moral status (artificial intelligence). The cases demonstrate that harm trajectories fall into recognisable types amenable to empirical investigation by the behavioural and systemic sciences.

The framework's dimensions apply to any entity capable of bearing harm on them, including non-human animals, ecosystems, and artificial systems whose moral status is uncertain. This is not an extension of the framework but a consequence of its foundational commitments. The plurality of the dimensions is presented not as a weakness to be overcome but as a structural feature of the moral domain that any adequate ethical framework must respect.

1. Thesis and Scope

Every decision is a harm profile, not a harm score.

Ethics is multi-dimensional, temporal, and irreducibly plural. The world's ethical traditions were never competing answers to the same question. They were partial instruments, each measuring different dimensions of the same underlying structure. What follows is an attempt to identify what those dimensions actually are, test whether they refer to something real, and work out what it means that they do not collapse into one.

The claim is not that ethics is subjective, or that all moral positions are equally valid. The claim is structural: the thing that ethics is about has more than one axis, those axes do not reduce to each other, and any action that reduces harm on one axis while increasing it on another is a *trade*, not a *solution*. The honest version of ethics makes the full structure of that trade visible, provides tools for eliminating options that are clearly worse, establishes floors below which no tradeoff is permitted, and narrows the space of genuinely contested decisions to those that should require judgment because they are genuinely hard.

What this document is

This is a diagnostic and constructive framework. It proposes a set of candidate dimensions along which real harm can occur, subjects each to tests for observability, independence, and measurability, and maps the world's major ethical traditions onto those dimensions to show what each tradition tracks, where it is strong, and where it is blind or silent. It develops a structured method for reasoning within the resulting multi-dimensional harm space and applies that method to nine historical cases to demonstrate both its diagnostic and its decision-guiding power.

The ambition is not to replace existing ethical theories but to identify the underlying variables they are, in different ways, responding to. If the framework is correct, the long-standing competition between traditions is not a contest with a winner. It is a division of labour that has never been recognised as such.

The framework is designed to connect to the empirical sciences. If the dimensions are real and observable, their dynamics over time are empirical phenomena, investigable by the methods of complexity science, systems dynamics, and the behavioural sciences. The philosophical contribution is the dimensional decomposition and the diagnostic standard. The empirical investigation of how those dimensions interact, propagate, and cascade is future work, but it is work for which tools already exist.

What this document is not

This is not a metaethical treatise. It does not attempt to settle whether moral properties are natural, non-natural, constructed, or illusory. It operates at a level that is compatible with several metaethical positions (discussed in Section 3) while being incompatible with others.

It is not an algorithm that produces a single answer from a set of inputs. It does, however, develop a structured method for moral reasoning that resolves most decisions through dominance-elimination, floor-constraints, and temporal analysis, and that narrows the genuinely contested cases to a smaller set than they initially appear to be. Within this framework, resolution is always partial, always temporal, and always revisable: the method produces structured narrowing, not terminal verdicts. The method is developed in Section 12 and applied to cases in Section 13.

It is not culturally neutral in its method, though it attempts to be culturally inclusive in its scope. The diagnostic standard it applies is broadly empiricist and naturalist in character. This is a substantive commitment, acknowledged as such, and defended on pragmatic rather than metaphysical grounds.

It is not restricted to human entities. The framework's dimensions apply to any entity or structure capable of bearing harm on them, including non-human animals, ecosystems, and potentially artificial systems whose moral status is uncertain. This is not an extension of the framework. It is a consequence of its foundational commitments, and it is tested in the case studies.

Structure of the argument

Section 2 argues that the standard framing of ethical theory as a competition between single-axis accounts is a category error, and locates this document's contribution relative to existing pluralist work (Ross, Berlin, Williams, Nagel). Section 3 sets out the methodological commitments and addresses the objection that empiricist criteria are inappropriate for a normative domain. Section 4 argues for harm as the central primitive and states the framework's normative presupposition. Sections 5 and 6 propose the candidate dimensions and introduce the temporal axis. Section 7 develops the incommensurability claim and the concept of the harm profile. Sections 8 and 9 map the traditions onto the dimensions and treat cross-cultural convergence as evidence. Sections 10 and 11 examine what does not survive the diagnostic standard and argue that plurality does not entail relativism. Section 12 develops the structured method for decision under pluralism. Section 13 applies the framework to nine cases. Section 14 states the open questions. Section 15 offers concluding observations.

2. Evaluative Monism and Its Alternatives

The monist inheritance

The dominant traditions in Western normative ethics share a structural assumption that is rarely examined as such: that the moral domain is ultimately organised around a single evaluative axis, and that the task of ethical theory is to identify which axis is fundamental. Classical utilitarianism proposes welfare. Kantian deontology proposes rational consistency and respect for autonomous agency. Natural law traditions propose conformity with a normatively loaded metaphysical order. Even many virtue-ethical accounts, despite their richer descriptive texture, tend to converge on eudaimonia as the one thing that human life is ultimately oriented toward.

This assumption, *evaluative monism*, has shaped not just individual theories but the structure of the debate between them. The inter-theoretic dispute in Western normative ethics is almost always framed as a competition: which single principle, correctly specified, captures what morality is really about? The expectation is that one tradition will eventually win, or that a sufficiently sophisticated synthesis will absorb the others.

The result is centuries of argument that presents itself as disagreement about answers but is, on closer inspection, disagreement about *which dimension of moral life to treat as fundamental*. When a utilitarian and a Kantian argue about whether it is permissible to lie to prevent a murder, they are not applying different methods to the same variable. They are tracking different variables (experiential harm in one case, the integrity of rational agency in the other) and each is treating the other's primary variable as either derivative or secondary. The disagreement is dimensional before it is substantive. If both dimensions are real and neither reduces to the other, the debate is not about who is right. It is about which form of harm to prioritise in this case, and the honest answer may be that both are real costs and the decision is a trade.

Evaluative monism also carries a temporal assumption that is less often noticed. Most single-axis frameworks evaluate actions at a point in time: what are the consequences now, or what does the rule require now, or what would a virtuous person do now? The possibility that the harm profile of a decision changes shape over time, that short-term and long-term profiles may point in different directions, and that the formation of the agent across a life is itself an ethical variable are all difficult to accommodate within a synchronic, single-axis framework. This temporal blindness is addressed in Section 6.

The non-Western counterpoint

It is worth noting that evaluative monism is substantially a Western inheritance. Many non-Western ethical traditions are already structurally plural in ways that the Western pluralist critics called for.

Confucian ethics tracks at least four dimensions simultaneously: self-cultivation (agency), the five relationships (relational integrity), ritual propriety (coordination), and harmony (stabil-

ity). It does not reduce these to a single axis or treat any one as derivative of the others. Buddhist ethics tracks the intersection of suffering (valence harm) and cognitive distortion (information harm) with unusual precision, and the Eightfold Path addresses harm-reduction across multiple dimensions simultaneously. Ubuntu and African ethical traditions treat relational integrity and coordination as co-constitutive of personhood rather than as competing values to be ranked. Indigenous relational ethics extends moral consideration beyond human entities to include land, ecology, and more-than-human kin, operating across relational, stability, and information dimensions without subordinating any to the others.

The significance of this for the present argument is twofold. First, it suggests that evaluative monism is a feature of a particular philosophical tradition, not a necessary structure of ethical thought as such. The assumption that ethics must reduce to one axis is an inheritance, not a discovery. Second, it means that the pluralist insight this document develops is not without precedent in the broader human ethical tradition; it has substantial precedent outside the Western canon, in traditions that were already tracking multiple dimensions before the Western pluralist critique began. The Western debate between monism and pluralism looks different when placed alongside traditions that were never monist to begin with.

The Western pluralist precedent

Within Western philosophy, the monist assumption has been challenged by a line of thinkers whose work this document builds on directly.

W.D. Ross argued in *The Right and the Good* (1930) that moral life is governed by a plurality of *prima facie* duties (fidelity, reparation, gratitude, justice, beneficence, self-improvement, non-maleficence) that cannot be reduced to a single principle and that may genuinely conflict. Isaiah Berlin extended the pluralist intuition to values themselves, arguing that liberty, equality, justice, and other fundamental goods are not merely difficult to combine but *incommensurable*: there is no common measure by which they can be ranked against each other. Bernard Williams pressed the case further, arguing that the reductive ambitions of systematic ethical theory, what he called ”the peculiar institution” of morality, distort moral experience by demanding a tidiness that the phenomena do not support. Thomas Nagel, in *Mortal Questions* and *The View from Nowhere*, identified five fundamentally different types of value (obligations, rights, utility, perfectionist ends, and personal commitments) and argued that their irreducibility to each other is not a deficiency in our understanding but a structural feature of the ethical landscape. More recently, moral particularists such as Jonathan Dancy have argued that the moral valence of any given feature, even a feature as apparently stable as ”causing pain”, shifts with context, resisting subsumption under fixed principles entirely.

So plurality in ethics is not a new idea, either within or beyond the Western tradition. What has been missing is a *diagnostic framework* that takes the pluralist insight seriously enough to ask the next set of questions: if moral life is genuinely multi-dimensional, what *exactly* are the

dimensions? Do they refer to real, observable features of entities and structures? Can they be specified with enough precision to support comparison across traditions? What is the relationship between the dimensions identified by independent traditions in different parts of the world? And what are the consequences, for decision-making, for institutional design, and for how we understand moral disagreement itself, of accepting that the dimensions do not reduce?

Ross named duties. Berlin named values. Williams named the distortion. Nagel named the types. The non-Western traditions named specific dimensions without requiring the monist detour. This document attempts to name the *variables* that all of these efforts are pointing at, to test whether those variables refer to something real, and to subject them to the kind of grounding tests that the philosophical tradition has generally not applied to its own primitives. The claim, developed in Section 4, is that these variables are dimensions of *harm*, understood as structural or functional degradation of something real.

The scalar temptation

The persistence of evaluative monism despite decades of pluralist critique, and despite the existence of non-monist traditions elsewhere in the world, deserves explanation. It is not simply an intellectual error. Monism offers something that pluralism, as usually stated, does not: *decision procedures*. If all moral considerations ultimately reduce to a single axis (that is, are reducible to a single value along which options can be ranked) then in any conflict between competing goods or competing harms, there is in principle a fact of the matter about which option is better, namely the one that scores higher on the fundamental axis. Pluralism, as traditionally presented, gives up this feature: it says the dimensions are irreducible, but then struggles to say what follows when they conflict.

This exchange, formal tractability purchased at the cost of dimensional accuracy, is the engine that keeps monist theories alive despite their well-documented shortcomings. Utilitarianism survives not because it handles agency, integrity, or relational obligation well (it does not), but because it produces a *ranking*. Kantianism survives not because it handles suffering, vulnerability, or systemic injustice well (it does not), but because it produces a *procedure*. The single-axis form is doing independent work, separate from the content of any particular theory. The decision-theoretic advantage of monism is real, and any pluralist framework that does not address it is incomplete.

This document takes the position that the single-axis form is itself a source of moral error. Every single-axis reduction hides the losses on the dimensions it dropped. Utilitarian calculation hides the cost to individual agency when an agent is overridden for aggregate benefit. Deontological constraint hides the cost in avoidable suffering when a rule is maintained despite catastrophic consequences. Virtue-ethical focus on character formation hides the cost of inaction in structural conflicts that no amount of personal cultivation will resolve. The losses are real. They are borne

by real entities. And the single-axis form makes them invisible, not by denying them but by providing no place in the accounting where they can appear.

The challenge, then, is not merely to reassert that ethics is plural (Ross, Berlin, Williams, and Nagel have done that, and the non-Western traditions never needed to). The challenge is to construct a pluralist framework that is *specific enough to be useful*: that names the dimensions precisely, tests them for reality and independence, addresses the decision-theoretic concern directly, and provides a structured method for reasoning within the resulting multi-dimensional space. Sections 5 through 7 of this document attempt the construction. Section 12 develops the method.

What if the plurality is the finding?

It is possible, and this document takes it to be likely, that the irreducibility of the dimensions is not a limitation of current ethical theory but a structural feature of the domain. The moral world may simply have more than one axis, in much the same way that the physical world has more than one fundamental force. If so, the search for the single master principle was not merely premature but misconceived. It is not a problem that will be solved by further theoretical work but a category error that should be retired.

This does not entail relativism. Multiple real dimensions is not the same as ”anything goes”. Some harms are more severe than others within a dimension. Some harm profiles are clearly worse than others across dimensions. Some tradeoffs are impermissible regardless of the gains they offer elsewhere, a point developed in Section 11. And the structured method developed in Section 12 resolves most decisions through dominance-elimination, floor-constraints, and temporal analysis, leaving a smaller residual of genuinely contested tradeoffs than the monist framing suggests. The plurality of the dimensions makes ethics more complex than any single-axis theory allows, but it does not make it intractable. It makes it honest.

The convergence evidence developed in Section 9 strengthens this position. If traditions separated by thousands of miles and centuries of time independently detect the same dimensions, and if those dimensions pass the diagnostic tests for observability, independence, and measurability, then the plurality is not a feature of our confusion about ethics. It is a feature of the territory that ethics is about.

3. Methodological Commitments

The diagnostic standard

This document applies a consistent set of tests to every candidate ethical dimension and every ethical tradition it considers. The tests are not themselves a moral theory. They are selection criteria for identifying which moral concepts are doing genuine explanatory work and which persist through inheritance rather than demonstrated explanatory power. The criteria apply symmetrically across traditions and across entity types: a Western philosophical concept receives no more default credibility than a non-Western one, and a concept applied to human entities is held to the same standard when applied to non-human entities or structures.

The standard has six components:

Refuse inherited categories. The words ”good,” ”right,” ”moral,” and ”ethical” are treated as pointers, not as explanations. If a tradition or dimension uses a term that cannot be unpacked into something more basic and observable, that term is flagged, not accepted as primitive. The question is always the same. What, specifically, does this term refer to? If the answer is circular or empty, the term is not doing work.

Look for the underlying observables. For each candidate dimension: does it refer to something that can be observed, at least in principle, in entities or structures? ”Suffering” refers to something observable. ”Moral wrongness” as a brute property does not. The threshold here is not that the dimension must be directly measurable in a laboratory sense, but that there must be some identifiable difference in the world between cases where the dimension is present and cases where it is absent. A dimension with no observable signature, however indirect, is not a candidate.

Test for independence. For each candidate dimension: is it reducible to some combination of the other dimensions, or does it track something they do not? Dimensions that collapse into others are retired. What survives is the minimal set. This is the criterion that distinguishes a genuine multi-dimensional account from a list. A list can include as many entries as one likes. A dimensional analysis requires that each axis carry information the others do not.

Test for measurability. At least partial measurability, not necessarily precise quantification, but enough practical traction that two observers could in principle agree on whether more or less of it is present. The standard is not the precision of physics. It is the rough but real comparability that we already rely on in practice: we can tell that one situation involves more suffering than another, that one agent has more autonomy than another, that one relationship is more damaged than another. If a proposed dimension cannot support even this level of comparative judgment, it is too vague to do ethical work.

Preserve plural structure where the evidence demands it. If two dimensions do not reduce to each other, do not force them into one. This is the methodological commitment that separates the framework from single-axis approaches. It is also the commitment most likely to

draw objections, since it means accepting that the output of ethical analysis is a profile rather than a ranking. The response to this objection is developed in Sections 7 and 12.

Treat convergence as evidence. When traditions separated by thousands of miles and centuries independently track the same dimension, that is evidence the dimension is robustly useful across contexts and tracks features of entities and structures stable enough to be detected independently. Convergence does not prove that the dimensions are mind-independent moral facts; it is compatible with several explanations, including shared cognitive architecture. But the combination of observability, independence from other dimensions, and cross-cultural convergence is a stronger case for the reality and usefulness of a dimension than any one of these criteria alone. The convergence argument is developed fully in Section 9.

Why these criteria?

The criteria above are, broadly, empiricist and naturalist in character. They privilege observability, measurability, and explanatory economy. The obvious objection is that ethics is a normative domain, not an empirical one, and that applying empirical selection criteria to normative concepts is a category mistake.

This objection deserves a direct response.

The claim is not that ethics *is* empirical science. The claim is that ethical concepts, like any other concepts, should be expected to refer to something. When a physicist proposes a new force, we ask: what would the world look like if this force existed, and what would it look like if it did not? When an ethicist proposes a new moral property, the parallel question is legitimate: what would the world look like if this property were present, and what would it look like if it were absent? If the answer is ”exactly the same, but with an additional non-natural property layered on top,” then the proposed property is not explanatory. It is explanatorily idle.

This does not require a commitment to full-blown moral naturalism. It does not require that moral properties be identical to natural properties, or that moral claims be reducible to empirical claims. What it requires is a minimal referential standard: if a moral concept is doing real work, it should be possible to say, at least roughly, what in the world it is tracking. Suffering tracks something. Agency tracks something. Relational integrity tracks something. ”Intrinsic moral wrongness,” understood as a brute non-natural property that supervenes on situations without being reducible to any feature of those situations, does not clearly track anything at all.

Metaethical compatibility

The diagnostic standard is compatible with a range of metaethical positions. It is compatible with naturalism, with certain forms of constructivism (those that treat moral properties as constructed from real features of entities and their interactions), and with pragmatist approaches that evaluate moral concepts by their practical usefulness. It is not compatible with strong non-naturalist real-

ism that treats moral properties as metaphysically primitive and unanalysable, nor with positions that treat moral terms as mere expressions of attitude with no referential content at all.

This is a substantive metaethical commitment, not a neutral methodological choice, and this document acknowledges it as such. The justification is pragmatic: concepts that meet the diagnostic standard have consistently proven more useful, more cross-culturally stable, and more resistant to motivated manipulation than concepts that do not. ”Harm” is harder to selectively redraw than ”the good.” ”Suffering” is harder to explain away than ”moral wrongness.” The history of ethics is, in part, a history of vague primitives being used to justify outcomes that more specific concepts would not support.

Connection to the empirical sciences

There is a further consequence of taking these criteria seriously. If ethical dimensions refer to real, observable features of entities and structures, then the dynamics of those features over time are empirical phenomena. They are investigable by the same methods that the behavioural and systemic sciences use to study other complex multi-dimensional systems: complexity science, systems dynamics, epidemiology, institutional economics, and conflict studies all model trajectories through multi-dimensional state spaces, looking for recurring patterns, tipping points, cascade effects, and feedback loops.

This is not an incidental observation. It is a structural consequence of the diagnostic standard. A framework whose dimensions are observable and at least partially measurable is a framework whose dynamics can, in principle, be studied empirically. The philosophical contribution of this document is the dimensional decomposition and the diagnostic standard. The empirical investigation of how those dimensions interact, propagate, and cascade over time is future work, but it is work for which tools already exist. The relationship between the framework and the empirical sciences is developed further in Section 12 and tested in the case studies in Section 13.

What the method cannot establish

Three limitations should be stated clearly.

First, the method can identify candidate dimensions and test them for reality, independence, and measurability. It cannot, by itself, establish that these dimensions *ought* to matter. The move from ”this is a real form of damage” to ”this damage ought to be prevented or minimised” is the familiar is-ought gap, and this document does not claim to have closed it. What it does claim is that the gap is narrower and more tractable when the ”is” side is well-specified. ”Suffering is real and this action increases it” is a better starting point for normative argument than ”this action is wrong because it violates the moral law.” Section 4 addresses this gap by stating the minimal normative presupposition the framework relies on.

Second, the method cannot determine in advance how many dimensions will survive. The framework proposes eight candidates and flags several as potentially reducible. The actual min-

imal basis may be smaller, or the reduction may not be achievable, leaving a larger set than is theoretically comfortable. The method commits to following the evidence on this question rather than forcing a predetermined answer.

Third, the method is stronger at identifying dimensions than at specifying how to navigate tradeoffs between them. This asymmetry is addressed in Section 12, which develops a structured method for reasoning within the multi-dimensional harm space. That method resolves most decisions through dominance-elimination, floor-constraints, and temporal analysis, but a residual set of genuinely contested tradeoffs remains where the framework makes the costs visible and the structure explicit without determining the outcome.

4. Harm as the Central Primitive

Why harm?

This framework centres on harm rather than on ”the good” or ”the right.” This choice requires justification, because it is not obvious and it is not innocent. It shapes everything that follows.

”The good” is underdefined. Ask three traditions what the good is and they will produce three answers that do not decompose the same way. Utilitarian accounts identify the good with welfare or preference satisfaction. Aristotelian accounts identify it with eudaimonia, a complex and contested notion of flourishing. Platonic and some religious accounts treat it as a metaphysical property apprehended by reason or revelation. In each case, ”the good” is used as if it were an explanatory primitive, but it does not explain. It labels. The explanatory work is being done by the more specific concepts underneath, and those concepts differ across traditions in ways that ”the good” as a unifying term obscures rather than illuminates.

”The right” is tradition-dependent. What counts as right action varies not only across cultures but across ethical theories within a single culture. For a consequentialist, the right action is the one that produces the best outcomes. For a Kantian, it is the one that conforms to the categorical imperative. For a virtue ethicist, it is the one that a person of practical wisdom would choose. ”The right” is the output of a theory, not a foundation for one. Building on it means importing an entire theoretical apparatus before the analysis has begun.

Harm is observable. Harm, understood as structural or functional degradation of something real, is cross-culturally recognised, empirically accessible, and does not require metaphysical commitments beyond ”some things can be damaged, and the damage is real.” A broken bone is harm. A severed relationship is harm. A collapsed institution is harm. A distorted belief system is harm. These are not metaphors. They are instances of real structures undergoing real degradation, and the degradation can be observed, at least in principle, across all of them.

Harm as the common genus across traditions

One further reason to centre harm: it is the concept that most reliably tracks what ethical traditions are actually responding to, across the widest range of cultural and philosophical contexts. Section 2 observed that the world’s ethical traditions track different dimensions. What is striking, when those traditions are examined through the lens of harm, is that each can be read as a response to a specific form of structural or functional degradation.

Utilitarianism responds to suffering and welfare-loss: the degradation of experiential states. Deontology responds to the violation of rational agency and the integrity of moral personhood: the degradation of the capacity to act from one’s own structure. Virtue ethics responds to character degradation: the erosion of stable dispositions necessary for a well-lived life. Care ethics responds to relational damage, neglect, and the failure of responsiveness to vulnerability: the degradation of bonds and care structures. Confucian ethics responds to social disharmony, role

failure, and the breakdown of cultivated relation: the degradation of the relational and coordinative fabric of social life. Buddhist ethics responds to suffering rooted in cognitive distortion and attachment: the degradation of internal models and the aversive states that distortion produces. Ubuntu and African ethical traditions respond to the breaking of communal personhood and reciprocal obligation: the degradation of the relational structures through which persons are constituted. Indigenous relational ethics responds to broken reciprocity with land, kin, and more-than-human community: the degradation of relational structures that extend beyond the human.

The word ”harm” is doing less work here than it first appears. It is not a theory. It is a *genus*. The real work is in the *species*: the specific dimensions along which harm occurs. Those are the subject of Section 5. But the observation that every tradition, when examined closely, is responding to some form of structural degradation is itself significant. It suggests that harm is not merely one possible primitive among several. It is the primitive that most reliably captures what all the traditions are pointing at, in their different ways, when they identify something as morally important.

What harm is not

Harm, as used in this framework, is not the same as discomfort, offence, or preference frustration.

A person who is told something true that they did not want to hear has not been harmed. A person whose preference for a particular outcome is not met has not necessarily been harmed. A community that encounters ideas it finds offensive has not, by that fact alone, been harmed. These experiences may be unpleasant, but unpleasantness is not the criterion. The criterion is *structural or functional degradation of something real*: an entity’s capacity to act, a relationship’s integrity, a structure’s stability, an organism’s welfare, a community’s capacity to coordinate.

This distinction matters because the concept of harm is routinely overextended in public discourse. If harm is allowed to include any experience of displeasure or any failure to get what one wants, then the concept loses its diagnostic power. Everything becomes harm, and harm becomes meaningless. The framework depends on harm retaining a specific referent: real damage to a real structure. The boundary between harm and mere discomfort is not always crisp, but the centre of each concept is clear, and the distinction between them is essential to the framework’s usefulness.

The normative presupposition

The framework as developed so far is descriptive: it identifies dimensions along which real harm can occur and proposes tests for their reality and independence. But the document’s practice, from Section 5 onward, moves from ”this is harm on dimension X” to ”this harm matters and should be tracked.” That move requires a normative commitment that should be stated explicitly rather than left implicit.

The commitment is this: *structural or functional degradation of entities capable of bearing it is pro tanto bad*. That is, harm on any of the proposed dimensions constitutes a reason, defeasible but real, for concern and for action. This is not derived from the framework's descriptive apparatus. It is a presupposition that the framework requires in order to do normative work.

The presupposition is minimal. It does not specify how much harm matters, how harms on different dimensions should be weighed against each other, or what obligations follow from the identification of harm. Those questions are addressed in later sections. What it does is close the gap between "this is real damage" and "this damage is morally relevant," a gap that the framework's diagnostic standard alone cannot bridge. The is-ought gap, as noted in Section 3, is not closed by this presupposition. It is narrowed to a single, clearly stated commitment that most participants in ethical discourse already accept, and that those who deny it (strict error theorists, for instance) would deny regardless of how the framework were constructed.

The negative-focus objection

The most immediate objection to centring harm is that it appears to reduce ethics to the avoidance of bad outcomes, ignoring positive goods: flourishing, excellence, beauty, love, creative achievement, spiritual development. If ethics is only about preventing harm, then a world in which nothing bad happens but nothing good happens either would count as morally perfect. That seems wrong.

The response is that the framework does not claim ethics is *only* about harm. It claims that harm is the most *diagnostically reliable* primitive available, and that positive goods, when examined carefully, tend to decompose into the *absence, prevention, or reversal of harm along specific dimensions*, or into states whose value is partly constituted by their vulnerability to harm.

Consider flourishing. A flourishing agent is one whose agency is intact (no agency harm), whose relationships are healthy (no relational harm), whose internal models are accurate (no information harm), whose social environment supports cooperation (no coordination harm), and who is not suffering (no valence harm). Flourishing is not a separate dimension added on top of harm-avoidance. It is what the harm dimensions look like when they are all in good order. This does not diminish the concept. It locates it more precisely.

Consider love. Love that is real and functioning involves relational integrity, mutual agency-respect, accurate mutual understanding, and the capacity to sustain these through perturbation. Each of these is the positive face of a harm dimension. Love is damaged when relational integrity is broken, when agency is overridden, when mutual models become distorted, or when the relationship cannot survive stress. The positive good and the corresponding harm dimension are two descriptions of the same axis, not two different things.

This is not negative utilitarianism. Negative utilitarianism holds that only the minimisation of suffering matters, and that positive welfare has no independent moral weight. The framework proposed here is structurally different. It holds that harm is the more *tractable* primitive for

identifying and comparing moral dimensions, not that harm is the only thing that matters. The positive states are real and valuable. They are also, for diagnostic purposes, better understood through the dimensions along which they can be damaged than through an underdefined notion of ”the good” that bundles them together without decomposing them.

The limits of the harm-centred approach

The decomposition of flourishing and love into harm dimensions is persuasive because these goods are constitutively relational and inter-personal: they decompose naturally into the proposed dimensions. But not all positive goods decompose so cleanly. Consider creative achievement: a composer writing a symphony, a mathematician proving a theorem. These are not merely states in which no harm is present on any dimension. There is a generative, additive quality, an exercise of capacity that goes beyond the mere non-degradation of that capacity. ”Exercising agency well” is not the same concept as ”absence of agency harm,” and in creative cases the former is doing the work.

The honest response is that the framework is stronger as a diagnostic for harm than as a complete description of positive goods, and this is a genuine limitation rather than a gap that can be papered over. The harm-centred approach captures most of what matters for decision-guidance, institutional design, and the identification of moral failure, which is where the framework’s practical contribution lies. But a complete account of moral life would need to say more about what constitutes a positively good life, a positively good society, and positively good creative and intellectual achievement than ”these are what the dimensions look like when nothing is going wrong.” The framework can say what damages a good life. It is less well-equipped to say what constitutes one. This limitation should be kept in view, particularly by readers from the Aristotelian or perfectionist traditions for whom positive goods are ethically central.

5. The Dimensions of Harm

Approach

What follows is a set of candidate axes along which real harm can occur. Each is subjected to the tests established in Section 3: observability, measurability, independence from the other candidates, and genuine explanatory work.

The format is deliberately uniform. Each candidate dimension is examined under the same headings so that comparisons across dimensions are straightforward and so that the reasoning behind retention or provisional flagging is visible. The goal is not to assert that these eight dimensions are the final answer, but to demonstrate a method: propose a candidate, test it, and retain or flag it based on what the tests reveal.

The tests are applied without presupposing a particular entity type. A dimension that tracks a real form of harm in human agents is equally a candidate for tracking that form of harm in non-human animals, ecosystems, or artificial systems, subject to the same observability and measurability requirements. Whether a given dimension applies to a given entity type is an empirical question, not a stipulation.

Where a dimension is retained, it is because it passed all four tests or passed three with a clear case for the fourth. Where a dimension is flagged as potentially reducible, the reasoning is stated. Where a candidate is rejected outright (Section 5.9), the grounds for rejection are given. The full mapping of ethical traditions onto these dimensions is developed in Section 8. The temporal axis along which harm profiles evolve is developed in Section 6.

5.1 Valence Harm

What it tracks: Suffering, aversive experiential states, negative phenomenal quality.

Observable? Yes, via behavioural indicators, neurophysiological correlates, self-report, and functional disruption. Entities that enter aversive states behave differently from those that do not: they withdraw, escape, protect, and signal distress. The observability of suffering is not in serious dispute, even if the precise nature of phenomenal experience remains contested.

Measurable? Partially. Intensity and duration can be roughly assessed. Cross-entity comparison is difficult, particularly across species or substrate boundaries. Precise quantification remains out of reach, but the difference between ”more suffering” and ”less suffering” is practically meaningful in most cases and is relied upon routinely in clinical, veterinary, and legal contexts.

Independent? Yes. An entity can suffer without losing agency, without relational damage, and without coordination breakdown. Pain in an otherwise functional, autonomous, well-connected agent is still harm. Conversely, serious harm on other dimensions (agency loss, relational erosion, information distortion) can occur without any suffering at all.

Traditions that track this primarily: Utilitarianism, Buddhist ethics, Mohism, aspects of care ethics.

What it misses when treated as the only dimension: Everything that is harmful but not experientially negative: slow erosion of agency, invisible relational decay, structural instability that has not yet produced suffering. The utilitarian tradition's long struggle with justice, integrity, and rights is a direct consequence of treating valence as the sole axis.

5.2 Agency Harm

What it tracks: Loss or degradation of the capacity to act from one's own structure: to form goals, select actions, and execute them without coercion or override.

Observable? Yes. Agents whose agency has been harmed show reduced behavioural range, increased dependency, inability to initiate, or compulsive compliance. The signs differ across entity types but the functional signature is recognisable: the agent is no longer steering. Coercion, manipulation, learned helplessness, and institutional disempowerment are all observable forms.

Measurable? Partially. Degrees of agency can be assessed through behavioural diversity, goal-persistence, resistance to override, and capacity for self-correction. Political philosophy and developmental psychology both work with rough but serviceable measures of autonomy.

Independent? Yes. An agent can retain full agency while suffering (a person in pain who is still choosing freely). An agent can lose agency without suffering (a person who has been subtly manipulated but feels fine). The Kantian tradition's insistence that respect for rational agency is not reducible to welfare is, in the terms of this framework, a correct observation about dimensional independence.

Traditions that track this primarily: Deontology, liberal political philosophy, existentialism, aspects of Hindu dharma traditions.

What it misses when treated as the only dimension: Harm that occurs through freely chosen action. An autonomous agent who freely destroys a relationship, or a society of autonomous agents who fail to coordinate, present moral problems that agency-centred accounts struggle to register.

5.3 Coherence Harm

What it tracks: Fragmentation of internal alignment: contradictions between beliefs, goals, values, and actions within a single agent or institution.

Observable? Yes. Incoherent agents produce contradictory outputs, unstable commitments, and decision paralysis or erratic oscillation. Cognitive dissonance, akrasia, and institutional self-contradiction are all observable forms.

Measurable? Likely, at least for formal systems. Logical consistency, goal alignment, and behavioural predictability from stated commitments can all be assessed. Measurement is messier in biological agents, but the direction is clear.

Independent? Partially. Coherence harm can occur without suffering (a person may hold contradictory beliefs comfortably) and without agency loss (a person may act freely but incoherently). However, severe coherence harm tends to degrade agency and stability over time, suggesting it may be partially derivative. It is retained provisionally as a candidate, with the acknowledgement that further analysis may reduce it to a combination of agency harm and stability harm.

Traditions that track this primarily: Deontology (universalisability as a coherence test), Stoicism, aspects of Confucian self-cultivation.

What it misses when treated as the only dimension: A perfectly coherent agent can be coherently monstrous. Internal consistency has no necessary connection to external harm. The most systematic perpetrators of atrocity have often been notable for the internal coherence of their belief systems.

5.4 Stability Harm

What it tracks: Degradation of persistence and robustness: the capacity of an agent, relationship, institution, or ecosystem to endure perturbation without collapse.

Observable? Yes. Fragile structures break under stress that robust ones absorb. The signs are visible in individuals (breakdown under pressure), relationships (inability to survive conflict), institutions (collapse under disruption), and ecosystems (tipping-point failure).

Measurable? Likely. Resilience metrics, recovery time, variance under perturbation, and failure thresholds can all be assessed across entity types. Ecology, engineering, and organisational science all work with stability measures, though they operationalise the concept differently.

Independent? Yes. A stable structure can be unjust (an oppressive regime that endures for centuries). An unstable structure can be otherwise well-functioning (a brilliant but fragile collaboration). Stability is not goodness, but instability is a real form of structural damage.

Traditions that track this primarily: Virtue ethics (character as stable disposition), Confucian ethics (social harmony), Daoist ethics (natural equilibrium), conservative political philosophy.

What it misses when treated as the only dimension: Stability can preserve harm. An unjust equilibrium is stable but not good. Stability as a lone criterion is conservative in the structural sense: it resists change, including change for the better. The most important moral transformations in history have involved the deliberate destabilisation of stable but harmful orders.

5.5 Relational Harm

What it tracks: Damage to the quality, integrity, or existence of relationships: bonds, dependencies, trust structures, care networks, reciprocal obligations.

Observable? Yes. Broken trust, severed bonds, neglected dependencies, betrayal, and abandonment are observable in behaviour, institutional function, and reported experience.

Measurable? Partially. Trust can be operationalised, and behavioural economics has done extensive work here. Relationship quality is harder to quantify but can be assessed through stability of mutual support, reciprocity patterns, and reported satisfaction. Care ethics and relational psychoanalysis both work with qualitative assessments of relational health that, while not numerically precise, support comparative judgment.

Independent? Yes. Relational harm can occur without suffering (a relationship slowly eroding without either party feeling acute pain), without agency loss (both parties remain free agents), and without coordination breakdown (they may still cooperate functionally while the relationship decays beneath the surface).

Traditions that track this primarily: Care ethics, Ubuntu and African ethical traditions, Confucian ethics, Indigenous relational ethics.

What it misses when treated as the only dimension: It can under-specify individual autonomy. A tradition that treats relational integrity as the master variable may struggle with cases where the right course of action is to break a relationship: to leave, to refuse, to prioritise self over bond. The tension between relational and agency harm is one of the most genuine and difficult tradeoffs in the framework.

5.6 Coordination Harm

What it tracks: Breakdown of the capacity for agents to cooperate predictably: loss of shared norms, trust infrastructure, institutional reliability, and cooperative equilibria.

Observable? Yes. Failed coordination is visible in defection, institutional collapse, norm erosion, inability to sustain collective action, and the breakdown of predictable interaction. The difference between a functioning institution and a dysfunctional one is not subtle.

Measurable? Likely. Game-theoretic metrics, institutional stability measures, norm compliance rates, and cooperation indices are all available. Real-world measurement is noisy but directionally clear.

Independent? Mostly. Coordination harm can occur without individual suffering (everyone is fine individually but cannot work together), without agency loss (everyone is free but cannot cooperate), and without relational damage in the intimate sense (strangers who cannot coordinate are not relationally damaged in the way that a betrayed friend is). However, coordination and relational harm overlap in small-group settings. Coordination harm is retained as independent because its large-scale, institutional form is clearly distinct from dyadic relational harm.

Traditions that track this primarily: Social contract theory, Mohism, Confucian ethics, game theory, institutional economics.

What it misses when treated as the only dimension: Coordination can be achieved through coercion. A system that coordinates perfectly by crushing dissent has solved coordination at the cost of agency and possibly valence. Coordination is a structural good, not a moral sufficient condition.

5.7 Information Harm

What it tracks: Distortion, corruption, or degradation of an agent's internal models: the accuracy with which it represents itself, others, and the world.

Observable? Yes. Agents operating on distorted models make systematically worse decisions, fail to predict outcomes, and are vulnerable to manipulation. Deception, propaganda, self-deception, and epistemic closure are all observable forms.

Measurable? Likely. Model accuracy can be assessed against outcomes. Calibration, prediction error, and susceptibility to manipulation are all measurable. Self-deception is harder but not invisible; it shows up in systematic prediction failure and in the gap between stated beliefs and revealed behaviour.

Independent? Yes. An agent can have distorted models without suffering (many comfortable delusions), without agency loss (a free agent acting on bad information), and without relational damage (a relationship sustained by mutual illusion). The Buddhist analysis of delusion (*moha*) as a root of suffering that is nevertheless distinct from suffering itself is, in these terms, a correct observation about dimensional independence.

Traditions that track this primarily: Buddhist ethics (delusion as a root cause of suffering), epistemology, scientific realism, parts of Islamic ethics (truthfulness as a core virtue).

What it misses when treated as the only dimension: Perfect information does not guarantee good action. An agent with a flawless model of the world can still choose cruelty. Information integrity is necessary for good decision-making but not sufficient for good outcomes.

5.8 Conflict Harm

What it tracks: Mutual interference, disruption, and zero-sum or negative-sum dynamics between agents: the degree to which agents damage each other through interaction.

Observable? Yes. Conflict is among the most visible of all social phenomena: war, competition, sabotage, antagonism, mutual obstruction.

Measurable? Likely. Conflict intensity, duration, destructiveness, and the degree of mutual damage can be assessed. The distinction between productive conflict (which generates information or corrects errors) and destructive conflict (which degrades all parties) is practically meaningful, though drawing the line between them in specific cases requires judgment.

Independent? Partially. Conflict often produces harm on other dimensions: suffering, relational damage, coordination breakdown, instability. But it is worth tracking separately because conflict *itself* can be the mechanism through which other harms are delivered. Reducing conflict without addressing the underlying dimensional harms may simply suppress the signal. Retained as a candidate but may be partially derivative.

Traditions that track this primarily: Daoist ethics (non-forcing, wu wei), Jain ethics (ahimsa, non-violence), peace studies, conflict resolution theory.

What it misses when treated as the only dimension: Conflict avoidance can be its own form of harm. Suppressing conflict to maintain false stability allows other harms to accumulate invisibly. Some conflict is necessary. The Daoist insight is about *unnecessary* forcing, not about the elimination of all friction.

5.9 Candidates That May Not Survive

Three concepts widely treated as ethically fundamental are, under the diagnostic standard of this framework, candidates for reduction rather than independent dimensions. Each tracks something real, but in each case the question is whether what it tracks is already captured by the dimensions above. The convergence evidence developed in Section 9 is relevant here: dimensions with strong cross-cultural convergence have a stronger claim to independence than those with weak or culturally specific convergence.

Fairness and symmetry. Impartial or symmetric treatment is widely invoked across ethical and political traditions, and the intuition that fairness matters is robust. The question is whether fairness constitutes an independent dimension of harm or whether it decomposes into harms already enumerated. Unfairness breaks cooperative equilibria (coordination harm). It overrides the interests of some agents for the benefit of others (agency harm). It can produce suffering in those who are treated unfairly (valence harm). It erodes trust in shared institutions (relational and coordination harm). If fairness-harm is fully accounted for by some combination of these dimensions, then it is not independent. It is retained provisionally, but under suspicion. The strongest case for its independence would be a demonstration that there exist cases of unfairness that produce no harm on any other dimension but are nevertheless genuinely harmful. Whether such cases exist is an open question.

Preference satisfaction. The degree to which an agent's goals are fulfilled is central to welfare economics and to preference-utilitarian accounts. But unmet preferences that do not produce suffering, agency loss, relational damage, or other dimensional harms are not clearly "harm" in the structural sense used here. A preference for vanilla over chocolate that is frustrated produces no degradation of any real structure. Preferences become ethically relevant when their frustration *produces harm on a real dimension*: suffering, agency loss, relational damage. Preference satisfaction is, under this analysis, a derivative concept rather than a primitive one. It is the surface at which deeper harms become visible, not an independent axis.

Complexity and richness. The degree of structured organisation in an entity or system tracks something real, and the loss of complexity is often a sign that something has gone wrong. But treating complexity-loss as harm in its own right introduces a bias toward complexity that is not obviously justified. A simple system that functions well is not harmed by being simple. A complex system that collapses into simplicity may have suffered stability harm, agency harm, or coordination harm, but the loss of complexity itself is the *symptom*, not the damage. Complexity is likely reducible to a combination of other dimensions and is not retained as an independent candidate.

The treatment of these candidates illustrates the diagnostic standard in its reductive mode. The standard does not merely accept proposed dimensions; it actively tests whether they are genuinely independent or whether they can be decomposed into more fundamental variables. This reductive discipline is what distinguishes the framework from a list of moral concerns and is applied to the framework's own proposals with the same rigour it applies to inherited concepts.

6. The Temporal Axis

Why time is not a modifier but an axis

The harm profile of a decision is not a snapshot. It extends across time, under uncertainty, and its shape can change as it does so.

This point is easy to state and difficult to take seriously in practice. Most ethical frameworks treat time as a background condition rather than as a structural feature of the moral landscape. A utilitarian calculates expected outcomes at some future point but treats the time-horizon as a parameter to be set, not as a dimension of the harm space itself. A Kantian asks whether a maxim can be universalised, but the test is synchronic: it asks about the logical structure of the act, not about what the act produces across a life or a generation. Even virtue ethics, which is more temporally sensitive than most traditions, tends to treat temporal development as the *context* for ethics rather than as a *dimension* of harm.

This framework takes the stronger position: time must be treated with the same seriousness as the harm dimensions themselves, though it is categorically different from them. Valence, agency, relational integrity, and the other dimensions proposed in Section 5 are axes *along which* harm occurs. Time is the axis *along which* harm profiles evolve. The harm dimensions tell us what kinds of damage are present. The temporal axis tells us how those kinds of damage develop, interact, compound, and sometimes reverse over time. A decision that reduces valence harm now may increase agency harm over years. A relationship that tolerates conflict in the short term may build stability over the long term. An institution that maximises coordination today through rigid control may collapse tomorrow when conditions change. These are not edge cases. They are the normal condition of moral life. Treating time as a background parameter rather than as a structural axis of the analysis is itself a source of moral error, because it allows the temporal dynamics of harm to remain invisible.

Temporal inversion

Short-term and long-term harm profiles can be inversely correlated. Protecting someone from suffering by removing their capacity to choose reduces valence harm at t_1 while increasing agency harm at t_{10}. Tolerating instability in a structure undergoing necessary change increases stability harm at t_1 while reducing it at t_{100}. Avoiding conflict to preserve relational comfort in the present may allow information harm and coordination harm to accumulate silently.

Temporal inversion is philosophically significant because it exposes a problem that single-axis frameworks cannot resolve internally. A utilitarian who evaluates only short-term welfare will endorse paternalistic interventions that a utilitarian evaluating long-term welfare would reject. The disagreement between them is not about the value of welfare. It is about the time-horizon over which welfare is assessed, and that choice is itself a moral decision that the utilitarian framework treats as exogenous. It has no internal resources for settling it. The multi-

dimensional framework addresses this directly: its structured method (developed in Section 12) requires that both short-term and long-term harm profiles be constructed and compared, making the temporal tradeoff explicit rather than allowing it to be resolved by an unexamined choice of time-horizon.

Intermediate timesteps

Harm trajectories do not jump from initial conditions to final outcomes. They pass through intermediate states, and the harm profile at each intermediate state may differ significantly from the profile at the origin or at any projected endpoint.

This observation has a practical consequence that the case studies in Section 13 develop in detail: escalating harm profiles often become legible at intermediate timesteps well before the most catastrophic outcomes materialise. A trajectory that begins with a negligible harm profile may, at intermediate stages, show escalating harm on multiple dimensions in ways that are detectable by anyone tracking more than one axis. The framework's prospective value lies substantially in this intermediate-stage legibility: not in predicting where a trajectory will end, but in recognising when a multi-dimensional harm profile is escalating in ways that warrant attention and potentially intervention.

The concept of intermediate-timestep analysis also clarifies the relationship between the framework and domain-specific empirical science. Scientists, economists, and other domain specialists detect signals within their own dimensions. The framework's contribution is to integrate signals from multiple dimensions into a single analytical frame at each timestep, making the full multi-dimensional profile visible where domain-specific assessments each see only their own dimension. This is conceptual organisation rather than signal detection, and it is a different and complementary kind of contribution.

Formation and cultivation

The virtue ethics, Confucian, and Buddhist traditions were tracking something real when they treated ethics as partly about *what kind of agent one becomes over time*. Character, cultivated disposition, trained perception, habituated response: these are not instantaneous. They are the result of harm profiles accumulated and navigated across a life. The question "what should I do right now?" is always entangled with the question "what kind of agent is this action building?"

This temporal dimension of ethics is systematically underweighted by traditions that treat moral life as a series of discrete decision-points. An ethics that evaluates each decision in isolation cannot register the harm of a trajectory. An agent who makes individually defensible decisions at every point but who is, over time, becoming less capable of relational integrity or less resistant to self-deception is undergoing a form of harm that only the temporal dimension reveals. The Aristotelian concept of *hexis* (stable disposition formed through repeated action)

and the Confucian emphasis on *xiuyang* (self-cultivation as ongoing practice) are both responses to this structural feature of moral life.

The implication for the framework is that harm profiles must be assessed not only across dimensions but across developmental trajectories. An action that produces no dimensional harm in isolation may, when repeated, produce cumulative harm that is severe. Conversely, an action that produces short-term harm on one dimension (the discomfort of honest self-examination, the instability of challenging a settled habit) may be constitutive of long-term harm-reduction on the same or other dimensions.

Intergenerational harm

Environmental law has built more serious legal and institutional architecture around intergenerational obligation than normative ethics has. The harm profiles of present decisions extend to people who do not yet exist, ecosystems that will bear the cost decades later, and coordination structures that future agents will inherit. Short-horizon optimisation is not neutral. It is a systematic bias toward the present that externalises harm onto the future.

The standard economic arguments for discounting future welfare (time-preference, uncertainty about whether future persons will exist or benefit, expected growth making future persons wealthier) are not without force. But the framework proposed here treats them differently than standard welfare economics does. Time-preference is a psychological fact about present agents, not a normative justification for weighting their interests more heavily. Uncertainty about the future is real, but it applies to *all* dimensions of the harm profile, not only to welfare, and uncertainty about whether harm will materialise is not the same as evidence that it will not. Expected growth is an empirical bet, not a moral principle, and it fails precisely in the cases that matter most: ecological tipping points, civilisational-scale coordination failures, and irreversible losses.

Derek Parfit's work on personal identity and future persons remains the most sustained philosophical engagement with these questions. His conclusion, that the interests of future persons are not less real for being temporally distant, is compatible with this framework's treatment of intergenerational harm. What the framework adds is the observation that intergenerational harm is not only a matter of welfare (Parfit's primary focus) but extends across all dimensions: future agents can inherit degraded coordination structures, depleted relational and institutional capital, distorted information environments, and reduced agency, not only reduced welfare.

No trajectory is ever closed

A consequence of taking the temporal axis seriously is that no harm profile is ever final. Every case, every decision, every trajectory remains open on every dimension. The harm profile at the current moment is different from the profile at any earlier point, and it will be different again.

This applies to cases that appear settled. A catastrophe that occurred decades ago is still producing effects: institutional responses (legal frameworks, norms, coordination structures) built

in reaction to it are still propagating harm-reduction. Intergenerational trauma is still propagating harm. The harm profile of the event at t_{100} is different from the profile at t_{30}, and a framework that treats the case as closed at any point is missing dimensions that are still in motion.

It also applies to cases of apparent success. A harm-reduction intervention whose benefits are still compounding is not a completed event but an ongoing trajectory. Its profile can change if the conditions that sustain the benefits are disrupted.

The implication is that the framework's temporal analysis is not a retrospective tool applied to finished cases. It is an ongoing analysis applied to trajectories that remain open. The case studies in Section 13 are presented in this spirit: not as settled verdicts on historical events but as readings of trajectories at their current point, subject to revision as the trajectories continue to evolve.

Uncertainty

Harm profiles across time are always partly uncertain. One cannot know with confidence what a decision's harm profile will look like in twenty years. This does not make the temporal axis less real. It makes it harder to navigate.

The precautionary instinct, that when uncertain about severe and irreversible harm one should err toward caution, is a reasonable heuristic for the temporal axis, but it is not a complete decision procedure. It addresses one specific problem (recklessness in the face of irreversibility) while leaving the deeper problem untouched: that temporal uncertainty makes the full harm profile of any decision partially unknowable in principle.

This is worth stating plainly rather than treating as a technical footnote. Decisions produce harm profiles across multiple dimensions across time, and temporal uncertainty is ineliminable. It follows that complete moral knowledge of one's own actions is structurally impossible. This is not a counsel of despair. It is a constraint on the kind of moral confidence that any framework, including this one, can legitimately claim. The appropriate response is not paralysis but epistemic honesty: act on the best available assessment of the harm profile, remain alert to evidence that the profile is changing, and do not mistake confidence about the short-term profile for knowledge of the long-term one.

Connection to the empirical sciences

The temporal dynamics described in this section are not unique to ethics. The behavioural and systemic sciences routinely study how multi-dimensional systems evolve over time: how feedback loops amplify or dampen initial perturbations, how tipping points produce non-linear transitions, how cascading failures propagate across coupled systems. If the harm dimensions proposed in this framework are real and observable, then these dynamics are empirically investigable using methods that already exist. The philosophical contribution is to identify the dimensions

and their moral significance. The investigation of how those dimensions interact and evolve over time is a question for empirical research, and the tools for conducting it are not hypothetical.

7. Incommensurability and the Harm Profile

The dimensions proposed in Section 5 are genuinely independent. Valence harm, agency harm, relational harm, coordination harm, and the rest do not reduce to each other. The temporal axis developed in Section 6 adds a further structural layer: harm profiles are not snapshots but trajectories that evolve, invert, compound, and sometimes reverse over time. It follows that any action that reduces harm on one dimension while increasing it on another, or that reduces harm now while increasing it later, is not a net gain or a net loss. It is a trade.

This is the central structural consequence of the framework, and it is worth developing carefully because it is the point at which the framework diverges most sharply from the dominant traditions.

Incommensurability

There is no universal exchange rate between dimensions. One cannot convert a unit of suffering into a unit of agency loss and determine which is "larger." The dimensions are incommensurable, not in the sense that they cannot be compared at all (Berlin's distinction between incommensurability and incomparability is relevant here), but in the sense that there is no single metric that converts all of them into a common currency.

This is the central deficiency of classical utilitarianism as a total moral framework: it assumes that such an exchange rate exists and that it is denominated in hedonic units. The history of the theory is largely the history of trying to make that assumption work, through preference-ordering, through welfare functions, through quality-adjusted life years. Each attempt has made the measurement more sophisticated while leaving the foundational problem untouched: that suffering and agency loss are different *kinds* of harm, not different *amounts* of the same harm.

The incommensurability claim should not be overstated. Within a single dimension, rough comparisons are often possible: more suffering is worse than less suffering, more coercion is worse than less coercion, more relational damage is worse than less relational damage. And across dimensions, some comparisons are clear enough in practice: a trivial loss on one dimension does not outweigh a catastrophic loss on another, regardless of which dimensions are involved. What is not available is a general-purpose conversion formula that works across all cases. The hard cases, where serious harms on different dimensions must be weighed against each other, are genuinely hard, and the framework does not pretend otherwise.

Harm profiles, not harm scores

Every decision produces a harm profile: a distribution of effects across the relevant dimensions, across time, under uncertainty. The profile is not a number. It is a shape, and the shape includes its temporal trajectory: how the distribution of harm across dimensions evolves as the consequences unfold. A decision whose harm profile looks favourable at t_1 may look very different at t_{20}, and the framework (following Section 6) requires that both readings be part of the analysis.

The appropriate question is never simply ”is this good or bad?” The appropriate question is: what is the shape of the harm this decision produces, across what dimensions, over what time, for which entities, and who bears the cost?

This reframing has a practical consequence that is easy to miss. When a decision is evaluated on a single axis, the losses on other axes are not merely underweighted. They are invisible. They do not appear in the accounting at all. A utilitarian cost-benefit analysis that shows a net welfare gain has no column for agency loss, relational damage, or coordination harm. These costs are real, they are borne by real entities, and the single-axis framework provides no place where they can be registered. The multi-dimensional framework does not eliminate the difficulty of weighing them, but it ensures they are at least *present* in the analysis.

The concept of a harm profile also clarifies a common source of moral disagreement. When two people look at the same decision and reach different moral conclusions, the disagreement may not be about values at all. It may be that they are looking at different dimensions of the same harm profile. One sees the suffering that is prevented. The other sees the agency that is overridden. Both are correct about the dimension they are tracking. This is, in miniature, the same pattern that the convergence argument in Section 9 identifies across ethical traditions: different instruments producing different readings of the same multi-dimensional phenomenon. Making the full profile visible does not guarantee agreement, but it transforms the nature of the disagreement from ”who has the right values?” to ”which dimensional costs are we willing to accept, and who bears them?”

Tradeoff transparency

The framework resolves most decisions through dominance-elimination and floor-constraints (developed fully in Sections 11 and 12). For the residual cases where serious harms on genuinely independent dimensions are in tension and no structural resolution is available, the framework does not collapse the tradeoff into a single answer. It insists that the costs on every dimension be named and attributed, so that those making the decision and those bearing the consequences can see what is actually being traded.

This residual transparency is sometimes characterised as a limitation relative to single-axis frameworks that claim to produce definitive answers in all cases. The characterisation is misleading. Single-axis frameworks achieve their apparent decisiveness by dropping dimensions from the analysis. The ”answer” they produce is definitive only because the costs on the dropped dimensions have been rendered invisible. A framework that makes all dimensional costs visible and resolves most cases structurally while requiring explicit judgment in the remainder is not weaker than one that resolves all cases by hiding costs. It is more honest.

What the framework does in contested cases is constrain the space of permissible tradeoffs. Some tradeoffs are not close calls: a decision that is better on every dimension than the alternative is simply better. Some tradeoffs are impermissible: the floors established in Section 11 specify

harms so severe that no gain on any other dimension can justify them. Between these bounds lies the genuinely contested territory where dimensional costs must be weighed against each other, and where the resolution is contextual, political, and always open to challenge. This is what the structured method in Section 12 calls structured narrowing rather than terminal verdicts: the method reduces the space of defensible options without collapsing it to a single point, and the residual judgment is informed by the structure rather than replacing it.

This is not moral paralysis

The concern that a multi-dimensional framework without a single-axis resolution must lead to paralysis is natural but mistaken. In practice, most decisions are not close calls across all dimensions. Many have harm profiles that are clearly better on most dimensions than the available alternatives. The difficult cases, where serious harms on different dimensions are in genuine tension, are precisely the cases that *should* be difficult, and any framework that makes them look easy is hiding costs rather than resolving them.

The framework supports a form of reasoning that single-axis theories do not: the identification of dominated options. An option that is worse on every dimension than some alternative can be eliminated regardless of how the dimensions are weighted. An option that is worse on most dimensions and only marginally better on one is a weak candidate that requires a strong justification for the marginal gain. Combined with floor-constraints and temporal analysis, this yields a structured method for narrowing the space of defensible decisions substantially, while preserving the openness that genuinely contested tradeoffs require. The full method is developed in Section 12.

8. What the Traditions Were Measuring

Reading the traditions as instruments

Every decision produces a harm profile: a multi-dimensional, temporally extended distribution of effects across the dimensions identified in Section 5, evolving along the temporal axis developed in Section 6. The world's ethical traditions are partial instruments, each tuned for sensitivity on some dimensions of that profile while being less responsive to others.

This is not a dismissive claim. Calling a tradition a "partial instrument" is not saying it is wrong. It is saying that it detects some dimensions of the harm space with high sensitivity while being blind or silent on others. A thermometer is not wrong about temperature because it cannot measure pressure. It is simply not the right instrument for that variable. The inter-theoretic debate in ethics has been, in large part, a debate between instruments that are each measuring different things, and recognising this transforms the debate from a competition into a division of labour.

What follows assesses each major tradition for which dimensions it tracks most sensitively, where its sensitivity drops off, what characteristic distortion results from treating its primary dimensions as the whole of ethics, and how it handles the temporal axis. Traditions that are temporally sensitive (tracking harm across developmental trajectories and intergenerational effects) have a structural advantage over those that evaluate actions at a single point in time. The case studies in Section 13 demonstrate concretely what happens when the dimensions a tradition drops from its analysis produce real harms that its framework renders invisible.

Single-axis and outcome-centred traditions

Utilitarianism is the tradition most finely tuned to valence harm. Its strength is that it takes suffering seriously and provides clear action-guidance through outcome comparison. No other Western tradition has done as much to insist that the experiential states of all affected parties enter the moral calculus. Its characteristic distortion is that it collapses all other dimensions into one, treating agency harm, relational harm, and coordination harm as mere contributors to the hedonic total. When these conflict with welfare maximisation, they are overridden, and the single-axis framework renders the override invisible: there is no column in the ledger where the cost to agency or relational integrity can appear. The long-standing objections to utilitarianism from the direction of justice (Rawls), integrity (Williams), and rights (Nozick) are, in the terms of this framework, objections to the dimensional collapse, not to the seriousness with which utilitarianism treats suffering. Its temporal sensitivity is limited: the standard framework evaluates outcomes at a chosen time-horizon without treating that choice as itself a moral variable, a problem that the temporal-inversion analysis in Section 6 makes explicit.

Mohism shares this outcome-centred structure but adds a stronger emphasis on coordination harm (anti-aggression, social order) and distributes concern impartially across all persons. It is

a more multi-dimensional consequentialism than Benthamite utilitarianism, and historically earlier. Its limitation is that impartial concern can flatten the special obligations and local relational structure that other traditions treat as ethically basic.

Constraint and rule-centred traditions

Deontology is the tradition most finely tuned to agency harm and coherence harm. The Kantian insistence that rational agents must never be treated merely as means is, in the terms of this framework, an insistence that agency harm cannot be traded away for gains on other dimensions. The categorical imperative's universalisability test is a coherence constraint: act only on maxims that could be willed as universal law without self-contradiction. These are genuine and important sensitivities. The characteristic distortion is detachment from consequences. A perfectly consistent, universalisable rule can produce catastrophic valence and relational harm, and the deontological framework provides no internal mechanism for registering that cost. The familiar trolley-problem literature is, at bottom, an exploration of this distortion. Its temporal sensitivity is limited: the universalisability test is synchronic, asking about the logical structure of the act rather than about the temporal trajectory of the harm profile it produces.

Parts of Islamic legal-ethical traditions and Hindu dharma traditions share this constraint structure while grounding it differently, in divine command and cosmic order respectively. The dimensional sensitivity is similar: both insist that certain actions are constrained regardless of outcomes. The justificatory architecture is different, and the difference matters for questions of legitimacy and cross-cultural applicability, but the dimensional profile is recognisably of the same type.

Character and cultivation traditions

Virtue ethics is the tradition most finely tuned to stability harm and agency formation across time. Its central question, what kind of agent is one becoming through the pattern of one's choices, is inherently temporal in a way that the decision-point frameworks of utilitarianism and deontology are not. Of the Western traditions, it is the one most naturally aligned with the temporal axis developed in Section 6. Its strength is that it takes character development, habituation, and the slow formation of practical wisdom seriously as moral variables. Its characteristic distortion is weak decision procedures: when stable character traits conflict, or when two people of practical wisdom disagree, the tradition offers *phronesis* as a placeholder rather than a resolution. This is not a trivial limitation. It means that in precisely the cases where guidance is most needed, the tradition defers to a faculty it cannot specify.

Confucian ethics combines character cultivation with relational and coordination sensitivity. It is one of the most multi-dimensional traditions in practice, tracking agency (self-cultivation), relational integrity (the five relationships, *wu lun*), social coordination (ritual propriety, *li*), and stability (harmony, *he*) simultaneously. Its temporal sensitivity is strong: the

emphasis on *xiuyang* (self-cultivation as ongoing practice) treats moral development as a lifelong trajectory rather than as a series of discrete decisions. Its limitation is role- and hierarchy-dependence: the relational structure it tracks is not egalitarian, and the roles it privileges can calcify coordination structures that ought to change. The tension between Confucian relational sensitivity and the egalitarian commitments of liberal ethics is, in these terms, a genuine dimensional tradeoff, not a simple case of one tradition being right and the other wrong.

Buddhist ethics is uniquely strong on the intersection of valence harm and information harm. Its core insight, that suffering is generated not only by external conditions but by distorted internal models (craving, *tanha*; aversion, *dosa*; delusion, *moha*), makes it the tradition most sensitive to the way information harm *produces* valence harm over time. The Eightfold Path can be read as a systematic programme for reducing harm across multiple dimensions simultaneously: right view and right intention address information harm, right speech and right action address relational and conflict harm, right livelihood addresses coordination harm, and right effort, right mindfulness, and right concentration address the formation of the agent over time. Its temporal sensitivity is inherent: the path is a practice sustained across a life, not a decision procedure applied at moments. Its limitation is measurement resistance: the internal states it tracks are real but difficult to assess from outside, and the tradition's tradeoff specification remains underdeveloped relative to the precision of its moral psychology.

Relational and communitarian traditions

Care ethics is the tradition most finely tuned to relational harm. It makes visible the dependencies, vulnerabilities, and care structures that more atomised frameworks miss. The work of Carol Gilligan, Nel Noddings, and Virginia Held established that a moral orientation centred on responsiveness to particular others, rather than on impartial principles, captures dimensions of moral life that principled frameworks systematically neglect. Its characteristic distortion is low formal structure: it can describe what matters without specifying how to decide when relational obligations conflict with each other or with obligations arising on other dimensions.

Ubuntu and African ethical traditions treat relational integrity and coordination as constitutive of personhood itself: *umuntu ngumuntu ngabantu*, a person is a person through other persons. This is a stronger claim than care ethics makes. It is not merely that relationships matter, but that the self is formed through relation and has no independent moral standing apart from its relational embedding. Its limitation is under-specification of individual autonomy when community demands conflict with personal integrity. The tension between Ubuntu's relational ontology and liberal individualism is one of the most substantive cross-cultural disagreements in the framework, and it cannot be resolved by declaring one side correct.

Indigenous relational ethics extends relational harm beyond human structures to include land, ecology, and more-than-human kin. It is the tradition most sensitive to harms that human-centred frameworks render invisible: the degradation of ecosystems, the severing of reciprocal

obligations to non-human entities, and the loss of place-based relational knowledge. It is also the tradition that most directly anticipates the framework's commitment to applying its dimensions to non-human entities and structures. Its limitation is the diversity of the traditions themselves. "Indigenous ethics" is not a single tradition but a vast family of place-specific practices, and generalising across them risks flattening precisely the contextual richness that gives them their diagnostic power.

Anti-reduction and context-sensitive traditions

Daoist ethics is tuned to conflict harm and the harm of over-intervention. Its central insight, that forced order can itself be a source of moral damage, is a direct sensitivity to the way that reducing harm on one dimension through control can increase harm on others. The concept of *wu wei* (non-forcing, or effortless action) is not passivity; it is an attunement to when intervention produces more harm than it prevents. Its limitation is indirectness: it is better at identifying what not to do than at guiding action in hard institutional contexts where inaction itself carries dimensional costs.

Jain ethics pushes harm-sensitivity further than almost any other tradition, extending non-violence (*ahimsa*) as broadly as possible while adding epistemic humility (*anekantavada*, many-sidedness). The Jain insistence that harm to any sentient being matters, combined with the recognition that one's own perspective is always partial, produces a framework that is both maximally harm-averse and maximally cautious about certainty. Its limitation is scalability: the demands of radical non-violence and epistemic humility are difficult to sustain under ordinary political and institutional conditions.

Moral particularism resists premature reduction. Its insight, developed most fully in Dancy's work, is that the ethical salience of any dimension shifts with context. A feature that is harmful in one situation may be benign or beneficial in another, and no fixed set of principles can capture this variability in advance. Its limitation is that it provides no stable decision framework. It is better understood as a corrective, a safeguard against the overfitting that any systematic framework (including this one) risks, rather than as a primary instrument in its own right.

Deflationary and meta-ethical traditions

Emotivism and error theory are not instruments for measuring harm. They are critiques of the assumption that moral language refers to anything at all. Emotivism, as developed by A.J. Ayer and C.L. Stevenson, holds that moral claims express attitudes rather than stating facts. Error theory, as developed by J.L. Mackie, holds that moral claims aim at truth but systematically fail because the moral properties they purport to refer to do not exist.

Both are powerful as diagnostics: they force traditions to show that their primitives refer to something real. Neither provides a reconstruction. Under this framework, their critique lands with full force against traditions that rely on undefined moral essences, but not against traditions

that track observable features of entities and structures. The dimensions proposed here are intended to survive the deflationary challenge. Suffering is not an attitude-projection. Agency loss is not a metaphysical fiction. Relational damage is not an error. If emotivism and error theory are correct that some moral language is non-referential, the appropriate response is to identify which moral concepts *do* refer to real features of the world and to build on those. That is what this framework attempts, and the normative presupposition stated in Section 4 (that structural degradation of entities capable of bearing it is pro tanto bad) is the minimal commitment that connects the referential success of these concepts to moral relevance.

The pattern across traditions

Several observations emerge from the survey as a whole.

Every substantive ethical tradition tracks at least one dimension with genuine sensitivity. None is simply wrong about what it detects. The diagnostic standard applied in Section 5 validates the dimensions that traditions have independently identified, even as it retires the theoretical superstructures some traditions have built around them.

The traditions that are most multi-dimensional in practice, tracking several dimensions simultaneously, tend to be the cultivation and relational traditions (Confucian, Buddhist, Ubuntu, Indigenous). The traditions that are most single-axis tend to be the Western systematic traditions (utilitarianism, deontology). This is consistent with the observation in Section 2 that evaluative monism is substantially a Western inheritance.

The traditions that are most temporally sensitive (virtue ethics, Confucian ethics, Buddhist ethics) are also the traditions that Section 6's temporal analysis most naturally extends. The traditions that are most synchronic (utilitarianism, deontology) are the ones for which the temporal axis represents the most significant structural addition.

The characteristic distortion of each tradition is predictable from its dimensional profile. A tradition that tracks only one dimension will systematically miss the harms produced on the dimensions it does not track. This is not a flaw in the tradition's detection apparatus. It is a structural consequence of treating a partial instrument as a complete one. The framework proposed in this document does not claim to transcend this problem. It claims to make the problem visible by holding all the instruments' readings in a single frame.

9. Convergence as Evidence

The argument from independent detection

Section 8 demonstrated that the world's major ethical traditions, despite profound differences in metaphysics, social context, and justificatory architecture, converge on a recognisable set of dimensions. Traditions that developed independently, separated by thousands of miles, centuries of time, and radically different starting assumptions, track the same forms of structural degradation. Valence harm is tracked across South Asia, East Asia, and Europe. Relational harm is tracked across sub-Saharan Africa, East Asia, Indigenous traditions, and Western care ethics. Coordination harm is tracked by social contract theory, Mohism, Confucian ethics, and Ubuntu. Information harm is tracked by Buddhist ethics, epistemology, and Islamic emphasis on truthfulness. Conflict harm is tracked by Daoism, Jainism, and peace traditions.

This convergence is not proof that the dimensions are real. But it is evidence, and it is evidence of a specific kind. When independent observers, working without knowledge of each other's results, detect the same variable, the most parsimonious explanation is that the variable exists and they are responding to it. The logic is analogous to independent measurement in the sciences: if two laboratories, using different instruments, report the same value for a physical constant, that is evidence that the constant is real rather than coincidence. The analogy is illustrative of the inferential structure (independent detection as evidence for the reality of what is detected) rather than a claim of full structural equivalence between ethical dimensions and physical constants. Ethical traditions are not laboratories, and dimensions of harm are not physical constants. But the inferential pattern is the same: independent detection of the same feature is better explained by the reality of the feature than by coordinated invention.

What counts as independent?

For the convergence argument to have force, the traditions must be genuinely independent, not merely superficially different branches of a shared intellectual lineage. The strongest cases are those where historical contact is minimal or absent.

Valence harm (suffering as morally central) is tracked by classical utilitarianism (developed in 18th-century Britain), Buddhist ethics (developed in 5th-century BCE India), and Mohism (developed in 5th-century BCE China). There is no evidence of mutual influence between Mozi and the Buddha, and neither had any contact with Bentham. The convergence on suffering as a primary moral variable across these three independent starting points is not a coincidence that requires no explanation.

Relational harm is tracked by care ethics (developed in late 20th-century North America), Ubuntu (rooted in sub-Saharan African traditions predating European contact), Confucian ethics (developed in classical China), and Indigenous relational ethics across multiple continents. These

traditions differ profoundly in their metaphysics, their social contexts, and their conceptions of personhood, yet all identify damage to relational structures as a primary moral concern.

The weaker cases are those where historical transmission is plausible or documented. The convergence between Kantian deontology and certain features of Islamic legal ethics on constraint structures could reflect shared Abrahamic or Hellenistic roots rather than independent detection. The convergence between Western virtue ethics and Confucian self-cultivation is somewhat stronger (the traditions developed independently) but both operate within literate, hierarchical, agrarian civilisations, which introduces a potential confound.

The argument does not require that every convergence be perfectly independent. It requires that enough of them are independent to make the hypothesis of shared detection more plausible than the hypothesis of coincidence or shared bias.

Convergence beyond the dimensions

The convergence evidence extends beyond the harm dimensions themselves. Two additional patterns are worth noting.

First, the importance of temporal sensitivity. Virtue ethics, Confucian ethics, and Buddhist ethics all independently treat moral development as a trajectory across a life rather than as a series of discrete decision-points. They converge not only on specific dimensions (agency formation, relational integrity, information accuracy) but on the structural insight that harm profiles evolve over time and that the temporal trajectory matters morally. This is convergence on the temporal axis developed in Section 6, and it strengthens the case for treating time as a structural feature of the harm space rather than as a background condition.

Second, the extension of moral consideration beyond human entities. Indigenous relational ethics across multiple continents, Jain ethics, and aspects of Buddhist ethics all independently extend harm-sensitivity to non-human entities and structures: ecosystems, animals, land, and more-than-human kin. This convergence supports the framework's commitment (stated in Section 1) to applying its dimensions to non-human entities as a consequence of its foundational commitments rather than as an ad hoc extension.

The cognitive-architecture objection

The strongest objection to the convergence argument is that cross-cultural convergence might reflect shared human cognitive architecture rather than shared moral reality. If human beings are, as a species, equipped with the same moral-psychological dispositions (an idea supported by some work in evolutionary moral psychology and by Haidt's moral foundations theory), then convergence on the same dimensions might reflect the structure of the human mind rather than the structure of the moral world.

This objection has force, and it cannot be fully resolved within the scope of this document. Three responses are available.

First, even if the dimensions reflect cognitive architecture, they may also track real features of the world. Human perceptual systems are products of evolutionary history, but they detect real properties (light, sound, temperature). The fact that our moral sensitivity has a biological basis does not entail that what it detects is unreal. If suffering is both a product of human (and animal) neurobiology and a real form of structural damage, then the cognitive-architecture explanation and the moral-reality explanation are not in competition. They are complementary.

Second, some of the dimensions tracked by convergent traditions are not plausible candidates for innate moral-psychological modules. Coordination harm, information harm, and stability harm are structural features of multi-agent systems that require a degree of abstraction to detect. If these dimensions converge across independent traditions, the explanation is less likely to be shared instinct and more likely to be shared exposure to the same structural features of social life. The convergence on temporal sensitivity and on non-human moral consideration, noted above, further weakens the cognitive-architecture explanation, since neither is a natural candidate for an innate moral module.

Third, the framework's diagnostic standard does not depend on settling the ultimate metaphysical question of whether the dimensions are mind-independent moral facts or species-wide cognitive projections. What it requires is that the dimensions refer to something observable and at least partially measurable, that they are independent of each other, and that they do genuine explanatory work. These criteria are met regardless of whether the dimensions are best understood as features of the world, features of human cognition, or (most plausibly) both.

To be precise about the claim: convergence is evidence that the dimensions are robustly useful across contexts and that they track features of entities and structures that are stable enough to be detected independently by traditions with no shared theoretical framework. This is a strong claim. It is not the stronger claim that the dimensions are mind-independent moral facts in the sense that physical constants are mind-independent physical facts. Whether the dimensions are ultimately best understood as features of an observer-independent moral landscape, as stable features of the interaction between human cognitive architecture and the social and physical world, or as some combination of both is a metaethical question that the framework does not need to settle in order to do its work. What matters for the framework's purposes is that the dimensions are real enough and stable enough to support cross-traditional comparison, diagnostic application, and the structured method developed in Section 12. The convergence evidence supports that claim.

Degrees of convergence

The pattern of convergence is informative about which dimensions are most robust.

The strongest convergence appears for valence harm, relational harm, coordination harm, and agency harm. Nearly every tradition surveyed tracks at least one of these as primary. These are

the dimensions for which the case for reality is strongest, because the evidence from independent detection is most extensive.

Strong convergence appears for information harm, stability harm, and conflict harm. These are tracked by multiple independent traditions, though sometimes as secondary concerns rather than primary ones. The case for their reality is solid but somewhat thinner, and in the case of conflict harm, the question of whether it is an independent dimension or a mechanism through which other harms are delivered remains open.

Weaker convergence appears for coherence harm (tracked primarily within Western philosophy, with some parallels in Confucian self-cultivation) and for fairness and symmetry (widely invoked but possibly reducible to coordination and agency harm). These may be real dimensions with culturally specific salience, or they may be derivative of more fundamental dimensions that happen to manifest differently across contexts. The framework flags them honestly rather than either asserting or denying their independence.

What the convergence evidence supports

The convergence pattern as a whole supports three claims that are central to the framework.

The first is that the dimensions proposed in Section 5 are not arbitrary. They were not selected by theoretical preference or constructed to fit a predetermined conclusion. They were identified by examining what independent traditions actually track, and the convergence across those traditions is evidence that the dimensions correspond to real features of entities and structures rather than to the idiosyncrasies of any particular philosophical lineage.

The second is that the framework's scope, applying the dimensions to non-human entities and extending the analysis across temporal trajectories, is not an ad hoc extension but is anticipated by traditions that independently arrived at both commitments. The convergence on these structural features strengthens the case for treating them as features of the moral domain rather than as optional additions.

The third is that the framework's practical apparatus, the structured method developed in Section 12, rests on dimensions whose cross-cultural robustness has been demonstrated. Dimensions with strong convergence evidence can be applied with greater confidence in the harm-profile analysis. Dimensions with weaker convergence evidence (coherence harm, conflict harm) are appropriately flagged as provisional. The grading of convergence evidence is not merely a theoretical exercise. It has practical consequences for how much weight a dimension should carry in the structured method.

10. What Does Not Survive the Diagnostic Standard

The diagnostic standard applied destructively

The diagnostic standard proposed in Section 3 is not only a tool for identifying dimensions that earn their place. It is also a tool for identifying concepts that do not. The same standard that validates certain dimensions through convergence evidence (Section 9) must also eliminate concepts that fail its tests, and this dual application is what gives the standard credibility. A standard that only confirms is not diagnostic. It is permissive.

What follows is not a dismissal of the traditions that rely on these concepts. It is an argument that certain foundational terms, as commonly deployed, do not meet the standard of observability, independence, and explanatory work that the framework requires. In several cases, the traditions that use these terms are doing real ethical work, but the work is being done by more specific concepts underneath, and the foundational term is obscuring rather than enabling the analysis.

Moral wrongness as a brute property

The claim that some actions are simply ”wrong”, not wrong because they produce harm on some dimension but wrong as an irreducible metaphysical fact, is a central commitment of strong non-naturalist moral realism. On this view, moral wrongness is a real property of actions, apprehended by moral intuition or rational insight, that cannot be reduced to any natural feature of the world.

Under the diagnostic standard, this position does not survive. Brute moral wrongness cannot be observed independently of the harm it is supposed to track. It cannot be measured. And it does no explanatory work that is not done better by specifying *which dimension of harm* is at stake. When someone says ”torture is wrong,” the moral force of the claim comes from the suffering involved (valence harm), the violation of the victim’s agency (agency harm), and the degradation of the relationship between torturer and victim (relational harm). The word ”wrong” adds a summary judgment but no additional information. Remove the dimensional content, and ”wrong” is empty. Retain the dimensional content, and ”wrong” is redundant.

This is not to say that the intuition of wrongness is unreal. The intuition is real, and it is often tracking something genuine. But what it tracks is harm on specific dimensions, and the brute-property framing obscures rather than illuminates this. ”It is wrong” is a conclusion. It is not a primitive. The normative presupposition stated in Section 4 (that structural degradation of entities capable of bearing it is pro tanto bad) does the normative work that ”wrongness” was supposed to do, but it does it with a stated commitment rather than an unexplained primitive.

”The good” as an unexplained unity

”The good” functions in much ethical discourse as if it were a natural kind: a single, unified property that ethical inquiry aims to identify and promote. Platonic, Aristotelian, and some religious traditions treat it as metaphysically basic.

Under the diagnostic standard, this unity does not hold. ”The good” decomposes into multiple independent dimensions (the absence or reversal of harm along the axes identified in Section 5), and the decomposition has more explanatory power than the unity. When two ethicists disagree about what the good requires in a given case, the disagreement almost always turns out to be a disagreement about which dimension of harm to prioritise, not a disagreement about a single property called ”the good.” The term bundles dissociable phenomena under a single label, and the label creates the illusion of a unified target where there are in fact multiple independent variables.

This parallels a broader pattern in the history of ideas. When a category that is treated as unified turns out, on closer inspection, to bundle dissociable components, the productive move is to retire the category and work with the components directly. The components are more precise, more measurable, and more useful for resolving disagreements than the false unity ever was.

Cosmic or divine order as an ethical foundation

Some traditions ground ethics in a cosmic order (dharma, Dao, divine command) that humans must align with. These frameworks are among the oldest and most influential in human history, and they deserve serious engagement rather than summary dismissal.

The diagnostic standard distinguishes two ways these frameworks can function. When the ”order” cashes out in terms of observable harm dimensions, the framework is doing real work. Dharma, understood as the pattern of situated obligations that minimises relational and coordination harm in a complex society, refers to something real and at least partially assessable. The Daoist concept of natural attunement, understood as sensitivity to when intervention produces more conflict harm than it prevents, refers to something observable. Divine command, when the commands track genuine dimensions of harm (prohibitions on killing, theft, and false witness all map onto specific harm dimensions), is operationally grounded even if its justificatory architecture is theological.

But when the order is left as a free-standing primitive that cannot be decomposed or observed, it fails the diagnostic standard. ”It is God’s will” or ”it is contrary to dharma,” stated without further specification, cannot be assessed for accuracy, cannot be measured, and cannot be used to resolve disagreements between people who hold different conceptions of the divine will or the cosmic order. The concept may motivate ethical behaviour, and the motivation may be powerful and genuine, but motivation is not the same as explanation. The diagnostic standard asks what a concept *refers to*, not what it *inspires*.

Purity and contamination

Moral concerns about purity, pollution, and contamination are widely present in human moral psychology. Jonathan Haidt's moral foundations theory includes sanctity/degradation as one of its core foundations, and anthropological work from Mary Douglas onward has documented the centrality of purity concerns across cultures.

The question for this framework is whether purity constitutes an independent dimension of harm or whether it is reducible to harms already enumerated. The evidence favours reduction. Disgust and contamination concerns, when examined in specific cases, typically function as proxies for other forms of harm: information harm (this substance or practice is dangerous; avoid it), conflict harm (this behaviour threatens group boundaries and social cohesion), or stability harm (this disrupts the established order and its associated roles). The moral charge of purity violations is real as a psychological phenomenon, but the underlying harm, when there is one, appears to be locatable on dimensions already in the framework.

This does not mean that purity concerns are always reducible without remainder. It is possible that there exist cases of purity violation that produce genuine moral distress without any corresponding harm on the other dimensions. If so, purity might warrant retention as an independent dimension. But the burden of proof lies with demonstrating such cases, and until that demonstration is made, parsimony favours treating purity as derivative rather than primitive.

The standard applied reflexively

The diagnostic standard must also be applied to the framework's own commitments. Section 4 acknowledges that the framework is stronger as a diagnostic for harm than as a complete description of positive goods, and that creative achievement and aesthetic experience do not decompose cleanly into harm-absence. Section 5 flags coherence harm and conflict harm as potentially reducible. Section 9 notes that convergence evidence is weaker for some dimensions than for others. These are not rhetorical concessions. They are applications of the same standard that eliminates brute wrongness and purity: concepts that do not pass the diagnostic tests are flagged regardless of whether they are convenient for the framework. A standard that exempts its own commitments from scrutiny is not a standard. It is advocacy.

What this section establishes

The diagnostic standard, applied destructively, produces a leaner ethical vocabulary. Brute moral wrongness is replaced by specified harm on identified dimensions. "The good" is replaced by the multi-dimensional harm profile and its positive face. Cosmic order is retained where it cashes out in observable dimensions and set aside where it does not. Purity is treated as derivative pending evidence to the contrary.

What survives is the set of concepts that refer to real, observable, at least partially measurable features of entities and structures. These are the concepts on which the framework builds.

The case studies in Section 13 demonstrate what happens when concepts that fail the diagnostic standard are used as the basis for decision-making: dimensions are dropped, harms are rendered invisible, and the costs are borne by entities whose interests the framework was not designed to track.

11. Plurality Is Not Relativism

The objection

The framework proposed in this document is multi-dimensional. It does not produce a single ranking of options. It holds that the dimensions of harm are incommensurable and that genuine tradeoffs between them cannot be resolved by formula. The most predictable objection is therefore: does this mean all moral positions are equally valid?

The short answer is no. The convergence evidence in Section 9 and the diagnostic standard applied in Section 10 already constrain what counts as a defensible moral position. Dimensions supported by strong cross-cultural convergence are not on the same footing as concepts that fail the diagnostic standard. Harm profiles grounded in observable dimensions are not equivalent to moral claims grounded in unexplained primitives. The plurality operates at the *dimensional* level, not at the *constraint* level. Multiple real dimensions of harm do not entail that all harm profiles are morally equivalent, any more than multiple physical dimensions entail that all spatial locations are equivalent. The structure of the space constrains what can be said about positions within it, even if it does not reduce to a single axis.

The longer answer requires distinguishing several mechanisms through which the framework supports substantive moral judgment despite its dimensional plurality.

Intra-dimensional ordering

Within each dimension, severity can be roughly ordered. Torture is more severe valence harm than mild discomfort. Total coercion is more severe agency harm than gentle persuasion. Systematic deception is more severe information harm than a white lie. The dimensions do not reduce to each other, but within each dimension the direction of ”more harm” and ”less harm” is clear enough to support comparative judgment in most cases.

This is not a trivial point. It means that the framework is not silent about severity. It can say, within any given dimension, that one state of affairs is worse than another. What it cannot do is convert severity on one dimension into severity on another. But that limitation is a consequence of the incommensurability argued for in Section 7, not a concession to relativism.

Floors

Some harms are severe enough that no gain on any other dimension justifies them. No coordination benefit justifies torture. No efficiency improvement justifies engineering mass suffering as invisible overhead. No stability gain justifies systematic dehumanisation.

These floors are not arbitrary. They correspond to what international human rights law calls non-derogable rights: constraints that hold regardless of emergency, expediency, or competing considerations. The Convention Against Torture permits no exceptions. The prohibition on

genocide admits no justification from aggregate welfare. These legal instruments reflect a cross-cultural judgment, built through decades of negotiation and contestation, that certain harms are so severe and so fundamental that no tradeoff calculus can reach them.

The framework adopts this structure. Floors are grounded in the normative presupposition stated in Section 4: that structural degradation of entities capable of bearing it is pro tanto bad. When that degradation reaches a threshold of severity on any dimension, the pro tanto reason becomes a decisive constraint. The philosophical basis for this is close to what Thomas Scanlon calls ”the priority of the worst off” and to what Judith Shklar calls ”the liberalism of fear”: the most urgent moral task is not the promotion of the best outcome but the prevention of the worst abuses.

The floors apply to all entities capable of bearing harm on the relevant dimension. A society that systematically degrades the agency of a subclass of its members is violating a floor regardless of whether its moral vocabulary recognises the subclass as rights-bearing. A system that produces severe and sustained valence harm to non-human animals is violating a floor regardless of whether the dominant framework treats animal suffering as morally relevant. The floors are set by the severity of the harm, not by the moral status the local tradition assigns to the entities bearing it.

The precise location of the floors is contestable, and this document does not claim to have settled it. What it does claim is that the existence of floors is a structural feature of any adequate moral framework, and that a pluralist framework can support them without requiring a single-axis reduction.

Dominance and near-dominance

Even without inter-dimensional commensurability, some harm profiles are clearly worse than others. A harm profile that is worse on every dimension than an available alternative is simply worse. This is the concept of dominance, borrowed from decision theory: an option that is dominated on all axes can be eliminated regardless of how the axes are weighted.

Near-dominance is also informative. A harm profile that is worse on most dimensions and marginally better on one is a weak candidate. It can be retained only if there is a strong justification for the marginal gain on the one dimension, and the burden of providing that justification falls on those who advocate for the near-dominated option.

Between dominance (clear) and genuine tradeoff (hard) lies a substantial range of cases where the multi-dimensional framework can say something substantive. The genuinely difficult cases, where serious harms on different dimensions are in real tension and no dominance relation holds, are the cases that *should* be difficult. Any framework that makes them look easy is concealing dimensional costs, not resolving them.

Temporal analysis

The temporal axis developed in Section 6 provides a further mechanism for substantive judgment. A harm profile whose short-term attractiveness masks long-term escalation across multiple dimensions is a weaker option than one whose profile is stable or improving over time. A decision that produces a favourable profile at t_1 but an escalating, multi-dimensional harm profile at intermediate timesteps is not equivalent to a decision whose profile remains favourable across the trajectory.

This is not a trivial addition to the anti-relativism argument. It means that the framework can distinguish between options that look equivalent at a single point in time but diverge over time. The temporal structure of the harm profile is itself a source of moral constraint, and it is available to the pluralist framework without requiring single-axis reduction. The case studies in Section 13 demonstrate this repeatedly: decisions that appeared defensible at the moment of action produced escalating multi-dimensional harm profiles whose temporal trajectories were legible at intermediate stages.

Differential quality of traditions

The framework supports a form of cross-traditional comparison that is distinct from cultural ranking. A tradition that tracks observable features of entities and structures with partial measurability is doing more genuine diagnostic work than a tradition that invokes undefined essences. This is not cultural chauvinism. The standard applies symmetrically to all traditions regardless of origin.

As Section 8 demonstrated, several non-Western traditions (Buddhist moral psychology, Confucian relational and cultivation ethics, Ubuntu, Jain non-violence, Indigenous relational ethics) outperform standard Western systematic theory on both groundedness and dimensional range. They track more dimensions simultaneously, they are more temporally sensitive, and they extend moral consideration more broadly. The diagnostic standard reveals this pattern; it does not impose a ranking from any particular cultural standpoint. The quality being assessed is instrumental: how many dimensions does the tradition detect, how reliably does it detect them, and how well does it handle the temporal axis?

The distinction from relativism

Moral relativism, in its strongest form, holds that moral claims are true only relative to a cultural or individual framework, and that there is no standpoint from which cross-framework judgments can be made. The framework proposed here rejects this. It holds that the dimensions of harm are real, that they can be detected across traditions, and that harm on any dimension is harm regardless of whether the local tradition recognises it as such.

What the framework shares with relativism is the rejection of a single universal ranking. What it does not share is the inference from plurality to equivalence. Berlin drew this distinction

clearly: to say that values are plural and sometimes incommensurable is not to say that all value-claims are equally defensible. Some positions are grounded in real features of the world. Others are not. Some harm profiles are clearly worse than others. Some floors are non-negotiable.

The framework's affirmative position is this: plurality constrains. Intra-dimensional ordering, floors, dominance, temporal analysis, and the diagnostic standard together produce a framework that is substantially more constraining than "anything goes" and substantially more honest than any single-axis reduction that achieves its apparent decisiveness by hiding dimensional costs. The structured method developed in Section 12 formalises these constraints into a decision procedure that resolves most cases and narrows the genuinely contested ones to those that should require judgment because they are genuinely hard.

12. Decision Under Pluralism

The decision-theoretic question

Section 11 established that plurality constrains: intra-dimensional ordering, floors, dominance, temporal analysis, and the diagnostic standard together produce substantive moral judgment without requiring single-axis reduction. The question now is how these constraints combine into a structured method for reasoning within the multi-dimensional harm space.

This is the question on which pluralist accounts have historically been weakest. Ross offered the judgment of the *phronimos*. Berlin offered tragic choice. Nagel offered fragmentation. None offered a method. This framework is in a position to do better, because it has built specific structural tools across the preceding sections.

The structured method

The method has five steps, each grounded in the preceding argument.

The first step is to construct the harm profile. For any decision, identify the effects across the relevant dimensions, for the relevant entities and structures, across the relevant time horizon, under the relevant uncertainty. The output is not a score but a profile: a distribution of harms and harm-reductions across dimensions and time. The question to be answered is not ”is this good or bad?” but ”what is the shape of the harm this decision produces, across what dimensions, over what time, for which entities, and who bears the cost?”

The second step is to eliminate dominated options. If one option is worse on every dimension than an available alternative, it can be eliminated regardless of how the dimensions are weighted. Near-dominated options (worse on most dimensions, marginally better on one) carry a burden of justification: the marginal gain must be shown to warrant the losses. This step resolves a substantial proportion of decisions without requiring any inter-dimensional commensurability.

The third step is to check against floors. If any remaining option violates a floor (produces harm so severe on any dimension that no gain elsewhere can justify it), it is excluded. The floors identified in Section 11, grounded in the normative presupposition of Section 4 and the non-derogable rights tradition of international human rights law, remove entire regions of the tradeoff space from consideration. The floors apply to all entities capable of bearing harm on the relevant dimension, not only to those the local framework recognises as rights-bearing.

The fourth step is to check the temporal trajectory. This is broader than checking for temporal inversions. It requires constructing the harm profile not only at the moment of decision but at intermediate timesteps and over the long-term trajectory, as developed in Section 6. Options that look attractive at t_1 but produce escalating harm across multiple dimensions at intermediate stages are weaker candidates than they initially appear. Options that impose short-term costs but reduce harm across multiple dimensions over time are stronger. The ”no trajectory is

ever closed" principle established in Section 6 applies: no temporal assessment is final, and the analysis should be revisited as the trajectory evolves.

The fifth step is to make the residual tradeoff explicit. After dominated options are eliminated, floors are applied, and temporal dynamics are accounted for, the remaining options represent genuinely contested tradeoffs: cases where serious harms on different dimensions are in real tension and no structural resolution is available. At this point the framework requires that the tradeoff be stated clearly, that the costs on each dimension be named, and that the question of who bears those costs be made visible. The resolution of such tradeoffs is contextual, political, and always open to challenge.

What this method achieves

The method is not an algorithm. It does not produce a single answer that follows mechanically from the inputs. It produces structured narrowing: eliminating dominated options, excluding floor-violating options, revealing temporal dynamics, and forcing the residual tradeoff into the open. Its value lies in reducing the space of defensible decisions to a substantially smaller set than it initially appears to be, while preserving the openness that genuinely contested tradeoffs require.

Most decisions that present themselves as irreducible tradeoffs turn out, under the full analysis, to be cases where one option is dominated, where a floor applies, where the temporal profile favours one side, or where the apparent conflict dissolves once the dimensions are specified precisely. The genuinely hard cases, where serious harms on independent dimensions are in real tension after all structural steps have been applied, are rarer than the current state of ethical debate suggests. They are also the cases that *should* require contextual judgment, because they are genuinely hard. Any framework that makes them look easy is concealing dimensional costs rather than resolving them.

The method is closer in structure to proportionality analysis in constitutional law or to clinical reasoning in medicine than to the deductive procedures of classical ethical theory. In each of these domains, a structured method narrows the space of defensible options without collapsing it to a single point, and the residual judgment is informed by the structure rather than replacing it.

The case studies in Section 13 are all presented as ongoing trajectories with genuine tradeoffs and open futures. None is presented as a case that the method settles definitively. This is not a failure of the method. It is a consequence of the framework's thesis that harm profiles are temporal objects and that no case is ever closed. The cases demonstrate what honest resolution looks like within the framework: always partial, always temporal, always revisable. A reader expecting terminal verdicts on settled cases is importing an expectation from single-axis frameworks that the framework explicitly rejects. The appearance of resolution in simpler cases is

often itself a form of dimensional blindness: the case looks closed because the dimensions that are still moving have been dropped from the analysis.

The direction of further development

The framework as presented here treats the dimensions as qualitative axes with rough ordinal comparisons within each. But there is nothing in its structure that prevents more formal development. The dimensions are proposed as real and at least partially measurable. If they are real and measurable, then formalisation is a natural next step, not a speculative addition.

Harm profiles could be represented as vectors across dimensions, with each component carrying a magnitude and a temporal trajectory. Multi-objective optimisation, a well-developed field in decision theory and operations research, already provides methods for working with vector-valued outcomes without requiring scalar reduction: Pareto frontiers, dominance relations, lexicographic orderings. These tools are structurally compatible with the framework's commitments. Temporal dynamics could be modelled as trajectories through the harm space rather than as static profiles: a decision does not produce a point in the harm space but a *path*, and dynamical systems modelling could capture how paths converge, diverge, or cascade.

Dimensional interactions could be mapped. The document has noted several cases where harm on one dimension produces harm on another over time: information harm producing valence harm (the Buddhist insight), agency harm producing relational harm, coordination harm producing stability harm. These are not merely correlations; they are causal pathways. A more developed version of the framework could specify these interaction effects formally, which would strengthen both the temporal analysis and the dominance-elimination step.

This direction does not require building new scientific disciplines. The behavioural and systemic sciences already study exactly these phenomena: how multi-dimensional systems evolve over time, how feedback loops amplify or dampen perturbations, how cascading failures propagate across coupled systems, how tipping points produce non-linear transitions. Complexity science, systems dynamics, epidemiology, institutional economics, and conflict studies all model trajectories through multi-dimensional state spaces and look for recurring patterns. What the framework adds to these disciplines is not a method but a normative structure: the identification of which dimensions carry moral significance, and the diagnostic standard for distinguishing real moral variables from inherited abstractions. The empirical sciences provide the dynamics. The framework provides the moral legibility. They are natural partners, and neither replaces the other.

This formalisation would require empirical work that has not yet been done: operationalising each dimension precisely enough to assign magnitudes, mapping the interaction effects, and validating the models against real cases. The framework has been designed to support this work. The dimensions are specified with observability and measurability criteria precisely so that the move from qualitative to quantitative analysis remains open.

What remains genuinely open

Two limitations should be stated clearly.

First, the framework does not resolve the residual tradeoffs that remain after all structural steps have been applied. Where serious harms on genuinely independent dimensions are in tension and no dominance relation holds, no floor is violated, and the temporal profiles do not disambiguate, the choice is contextual and political. The framework insists that this choice be made transparently, with the costs on each dimension named and attributed. It does not make the choice itself.

Second, the framework does not specify how to set the floors. It argues that floors exist and that they are grounded in the severity of harm on its own dimension and in the normative presupposition of Section 4, but the precise location of the boundary between ”severe enough to be a floor” and ”severe but tradeable” is itself a normative judgment that the framework constrains but does not determine. The non-derogable rights tradition provides the strongest existing guidance, but that tradition is itself the product of political negotiation, not of derivation from first principles.

13. The Framework Applied: Nine Cases

Purpose

The preceding sections have described the framework's structure. This section demonstrates it in use, applied to nine cases, each chosen because its trajectory has developed over sufficient time to reveal multi-dimensional dynamics and because it demonstrates a different structural feature of the analysis.

Each case applies the full method developed in Section 12. The harm profile is constructed across all eight dimensions at multiple timesteps, so that the temporal trajectory of the profile is visible. No case is treated as settled or concluded. If the framework is correct that harm profiles are temporal objects, then every case is an ongoing trajectory, and the analysis presented here is a reading at the current point in that trajectory rather than a final verdict. The cases are not hypotheticals. They are real, researched, and contested, which makes them a harder and more informative test than any thought experiment.

Case 1: The rejection of Hitler from the Vienna Academy of Fine Arts (1907)

What the framework reveals: trajectory harm and temporal inversion at civilisational scale.

In 1907 and again in 1908, the Vienna Academy of Fine Arts rejected the application of a young Adolf Hitler. The case demonstrates that a near-zero harm profile can precede a maximal one, that the trajectory between them passes through intermediate stages where the escalation is legible, and that the effects of the catastrophe are still propagating more than a century later.

The harm profile at t_1 (1907, the moment of rejection):

Valence harm: minor. The disappointment of a rejected applicant. Agency harm: minor. A path was closed, though others remained open. The applicant's capacity to act and choose was not significantly degraded. Coherence harm: negligible. No internal contradictions or fragmentations were produced by the rejection itself. Stability harm: negligible. The applicant's life circumstances were disrupted but not destabilised in any severe sense. Relational harm: negligible. No significant relationships were damaged by the rejection. Coordination harm: none. The rejection had no effect on social coordination structures. Information harm: none. No models were distorted. Conflict harm: none.

By any evaluation at t_1, whether single-axis or multi-dimensional, this is a trivially low-harm event.

The harm profile at t_{15-30} (1920s-1930s, intermediate stages):

The trajectory did not jump from negligible to catastrophic in a single step. It passed through intermediate states, including homelessness and radicalisation in Vienna, the experience of the First World War, the Beer Hall Putsch, imprisonment, ideological consolidation, and institutional

capture. At these intermediate stages, the harm profile was escalating visibly across multiple dimensions.

Valence harm: increasing. Political violence, intimidation, and the early persecution of targeted groups were producing measurable suffering. Agency harm: escalating. Political coercion, the suppression of opposition, the abolition of independent institutions, and the progressive destruction of individual autonomy were all detectable. Coherence harm: increasing. An ideology that required systematic self-deception and doublethink from its adherents was producing institutional and individual incoherence. Stability harm: increasing. Democratic institutions were being subverted and dismantled. The political structures of the Weimar Republic were being destabilised deliberately. Relational harm: escalating. Systematic scapegoating and social division were tearing communities apart along racial, political, and religious lines. Coordination harm: escalating. Institutional subversion, the destruction of independent media, the capture of the judiciary, and the replacement of cooperative norms with coercive ones were degrading the coordination fabric of German society. Information harm: escalating. A propaganda apparatus was being constructed that would make systematic deception a governing principle, distorting the internal models of an entire population. Conflict harm: escalating. Political violence, paramilitary activity, and the suppression of dissent were all increasing.

A multi-dimensional analysis at this intermediate stage would have detected escalating harm across every dimension. The signal was present. The framework's prospective value in cases like this lies not in predicting trajectories from their negligible origins but in recognising escalating multi-dimensional harm profiles at intermediate stages, where intervention is still possible.

The harm profile at t_{30-38} (1939-1945, the catastrophic phase):

Valence harm: maximal. The suffering and death of tens of millions, including systematic extermination, mass displacement, famine, and the psychological devastation of survivors. Agency harm: maximal. Totalitarian destruction of individual autonomy across an entire continent, including the abolition of political participation, forced labour, and the reduction of entire populations to objects of state policy. Coherence harm: severe. The ideology required its adherents to maintain systematic self-deception on an institutional and individual scale. Stability harm: catastrophic. Continent-wide physical destruction and institutional collapse, the dismantling of political, legal, and cultural structures that had taken centuries to build. Relational harm: catastrophic. Communities, families, and social bonds torn apart systematically through deportation, persecution, betrayal, and genocide. Coordination harm: catastrophic. The international order built since the Congress of Vienna collapsed; the cooperative structures of European civilisation were shattered. Information harm: severe. A propaganda state that made systematic deception a governing principle, distorting the internal models of entire populations across occupied Europe. Conflict harm: maximal. Global war, with destruction extending across continents and oceans.

No single-axis framework captures this profile. The utilitarian sees the suffering. The Kantian sees the agency violations. The virtue ethicist sees the character degradation. Each captures one dimension. The multi-dimensional framework sees the full profile and, critically, it sees the temporal structure: a near-zero profile at t_1, an escalating profile at t_{15-30} that was legible to anyone tracking multiple dimensions, and a maximal profile at t_{30-38}.

The harm profile at t_{120} (the present, ongoing effects):

The trajectory did not end in 1945. The post-war response to the catastrophe produced its own multi-dimensional effects.

Valence harm: substantially reduced from the catastrophic phase, but intergenerational trauma continues to propagate across affected populations. Agency harm: substantially reduced. The post-war reconstruction of democratic institutions restored agency at scale, but the geopolitical structures of the Cold War (themselves a consequence of the catastrophe) produced agency harm across Eastern Europe for decades. Coherence harm: reduced but not eliminated. The processing of the catastrophe remains incomplete in many societies, and the political instrumentalisation of the period produces its own forms of institutional incoherence. Stability harm: substantially reduced. The post-war order (the United Nations, the European integration project, NATO) produced stability structures that are still operative. But the displacement and statelessness of populations whose consequences persist to the present represent ongoing stability harm. Relational harm: partially repaired through institutional responses (reconciliation processes, memorial culture, reparations programmes) but intergenerational relational damage persists. Coordination harm: substantially reduced. The United Nations, the Universal Declaration of Human Rights, the European Convention on Human Rights, the Nuremberg principles, and the Convention Against Torture are coordination structures built specifically in response to the catastrophe. These represent ongoing coordination-harm-reduction that is still operative. Information harm: substantially reduced through historical documentation, education, and memorial institutions, though revisionism and denial continue to produce information harm. Conflict harm: substantially reduced from the catastrophic phase, but the Cold War geopolitical structures that emerged from the catastrophe produced decades of proxy conflict.

The harm profile at t_{120} is different from the profile at t_1, at t_{15-30}, at t_{30-38}, or at t_{80}. The institutional responses to the catastrophe are themselves ongoing trajectories of harm-reduction. The continuing harms are themselves ongoing trajectories that have not concluded. No case is ever closed.

Case 2: Henrietta Lacks and the HeLa cell line (1951)

What the framework reveals: genuine multi-dimensional tradeoff with no clean resolution, partial repair over time, and compounding cascading benefits.

In 1951, cells were taken from Henrietta Lacks, a Black woman being treated for cervical cancer at Johns Hopkins Hospital, without her knowledge or consent. Those cells, the HeLa

line, became the most widely used human cell line in biological research. They were essential to the development of the polio vaccine, contributed to advances in cancer treatment, virology, genetics, and countless other fields, and continue to be used in laboratories worldwide. The case demonstrates that a genuine multi-dimensional tradeoff, where real benefits and real harms coexist on independent dimensions, can evolve over time as harms are partially repaired and benefits continue to compound.

The harm profile at t_1 (1951, the taking of the cells):

Valence harm: to Lacks herself, severe (she was dying of cancer; the cell extraction added no additional suffering, but her illness and death were sources of profound suffering). To the broader population, no immediate effect. Agency harm: severe. Lacks was never asked. She never consented. Her autonomous right to decide what happened to her own body was not merely overridden; it was never acknowledged as relevant. The racial context is inseparable from this: a Black woman in a segregated hospital in 1951 was not treated as an agent whose consent mattered. Coherence harm: negligible at the individual level, though the institutional practice of taking tissue without consent reflected an incoherence between the medical establishment's stated values and its actual behaviour. Stability harm: negligible. No structures were destabilised by the act itself. Relational harm: present but not yet visible. The act would later damage the relationship between the Lacks family and the medical establishment, but at t_1 the family did not know the cells had been taken. Coordination harm: negligible. The act did not disrupt any coordination structures. Information harm: present. The family was not informed about what had been done. The information asymmetry between the institution and the patient was itself a form of structural harm. Conflict harm: negligible at t_1.

The harm profile at t_{20-60} (1970s-2010, disclosure and compounding benefits):

The harm side of the profile became visible and intensified on some dimensions during this period. The benefit side began to compound at scale.

Valence harm: massively reduced at population scale. The medical advances enabled by the HeLa line alleviated suffering for millions of people across decades. The polio vaccine alone prevented enormous suffering. For the Lacks family, the discovery of what had been done produced its own distress. Agency harm: reduced at population scale (people who would have died or been severely disabled retained their capacity to act, choose, and participate in their own lives). For the Lacks family, agency harm persisted: they had no control over the use of the cells, no voice in the research, and no share in the commercial value generated. Coherence harm: increasing at the institutional level, as the gap between the medical establishment's stated commitment to patient rights and its history of taking tissue without consent became publicly visible. Stability harm: negligible. The case did not destabilise any structures, though it contributed to the destabilisation of the previous norm of unquestioned tissue use. Relational harm: compounding for the Lacks family. They were not informed about the cell line's existence for over twenty years.

When they did learn, the institutional response was inadequate, and the commercial use of the cells without any benefit to the family added a further layer of relational damage. Coordination harm: reduced at population scale (the polio vaccine enabled public health infrastructure and social cooperation that would have been impossible with endemic polio). For the medical establishment, the case began to produce coordination-harm-reduction through the reshaping of informed consent norms. Information harm: partially reduced for the Lacks family when they learned the truth, and further reduced through Rebecca Skloot's 2010 book, which brought the case to public attention. At population scale, information harm was massively reduced: the HeLa line advanced biological knowledge profoundly, correcting distorted models of disease, viral behaviour, and cellular function. Conflict harm: increasing. The conflict between the Lacks family and the medical establishment intensified as the case became public.

The harm profile at t_{75} (the present, ongoing effects):

Valence harm: continuing to reduce at population scale. The HeLa line is still in active use. COVID-19 vaccine development used HeLa cells. Ongoing research in cancer treatment, gene therapy, drug testing, and virology continues to produce new harm-reduction. Each new medical advance generates its own downstream cascade: a person whose life is saved retains their agency, participates in relationships, contributes to institutions and communities, and produces further effects across dimensions. The benefit side is not a fixed quantity delivered in the 1950s. It is a compounding multi-dimensional cascade that is still growing. Agency harm: partially repaired for the Lacks family through the NIH's 2013 agreement, which gave them a degree of control over access to the HeLa genome. The family now has a voice. At population scale, agency-harm-reduction continues to compound. Coherence harm: partially addressed through institutional acknowledgment and policy reform, though the tension between historical practice and current values remains unresolved. Stability harm: negligible. Relational harm: partially repaired. The relationship between the Lacks family and the medical establishment has been partially addressed through institutional responses, though the original violation cannot be undone. Coordination harm: reduced at multiple scales. The case has reshaped informed consent frameworks, bioethics policy, and research ethics worldwide. These institutional changes constitute ongoing coordination-harm-reduction that extends far beyond the Lacks family. Information harm: substantially reduced. The case is now widely known, documented, and taught. The information asymmetry that characterised the original act has been largely corrected. Conflict harm: reduced. The most intense period of conflict between the family and the institutions has been partially resolved through the NIH agreement and public recognition, though the underlying tensions have not disappeared entirely.

No single-axis framework resolves this case. A utilitarian calculus that compares the aggregate welfare gained against the suffering of one woman and her family will produce a clear answer: the benefits outweigh the costs by orders of magnitude. But that answer collapses the

multi-dimensional benefits into a single welfare figure and hides the agency harm, the relational harm, and the information harm entirely. A Kantian analysis correctly identifies the agency violation but has difficulty accounting for the genuine and multi-dimensional benefits that resulted.

The multi-dimensional framework produces a different and more honest output. It makes the tradeoff structure visible in its full complexity. The benefits are multi-dimensional and still compounding. The harms were also multi-dimensional but have been partially repaired over time. Neither set cancels the other. The case is a genuine tradeoff, not a puzzle with a hidden correct answer, and the appropriate response is to insist that all dimensions remain visible: that the benefits be acknowledged *and* that the original violations be acknowledged as real costs that were borne by real people. The trajectory remains open.

Case 3: Daniel Ellsberg and the Pentagon Papers (1971)

What the framework reveals: information-harm correction with short-term costs and long-term gains, assessed against the escalating harm profile of the status quo it disrupted.

In 1971, Daniel Ellsberg, a military analyst, leaked a classified Department of Defense study of United States decision-making in Vietnam to the New York Times and other newspapers. The study revealed that the government had systematically deceived the public about the progress and prospects of the war, and that senior officials had privately assessed the war as unwinnable while publicly insisting it could be won.

Assessing this case requires constructing two profiles: the harm profile of the act itself and the harm profile of the status quo the act disrupted. The leak cannot be evaluated in isolation. It must be evaluated against what was already happening on all dimensions.

The harm profile of the status quo at the time of the leak (the escalating baseline):

By 1971, the status quo was producing escalating harm across multiple dimensions through intermediate stages that had been legible for years. The deception did not begin at full scale. In the early 1960s, there were optimistic assessments that may have been genuinely held. By the mid-1960s, private doubts coexisted with public confidence. By the late 1960s, the gap between what the government knew and what it told the public had become systematic and sustained deliberately.

Valence harm: severe and escalating. Casualties were mounting in a war assessed internally as unwinnable. Suffering extended across military personnel, Vietnamese civilians, and the families of both. Agency harm: severe. Conscription expanded to sustain a policy known internally to be failing, overriding the agency of those compelled to serve. Coherence harm: severe at the institutional level. The government was maintaining a systematic gap between its internal assessments and its public statements, a form of institutional incoherence that was corroding decision-making. Stability harm: increasing. The war was destabilising Southeast Asia and producing domestic political instability in the United States. Relational harm: severe. The war was destroying communities and families on both sides. The deception was eroding the rela-

tionship between the government and the public. Coordination harm: increasing. International coordination structures were being strained by the war and by the credibility gap. Information harm: severe and escalating. The public's model of the war and of their government's honesty was systematically distorted. Conflict harm: severe. The war itself, and the domestic political conflict it generated, were producing harm at scale.

The harm profile at t_1 (1971, the act of leaking):

Valence harm: mixed. The immediate emotional and professional consequences for Ellsberg were severe (criminal prosecution, personal upheaval). No direct physical suffering resulted from the leak itself. Agency harm: mixed. Ellsberg exercised his own agency at great personal risk, acting from his own moral structure rather than complying with an institution he believed was acting wrongly. But the leak constrained the agency of others who had operated within the classification system in good faith. Coherence harm: reduced. The leak exposed the institutional incoherence between private assessment and public statement, beginning the process of correcting it. Stability harm: increased in the short term. The leak destabilised the relationship between the press, the public, and the government, and contributed to a broader crisis of institutional legitimacy. Relational harm: increased. Ellsberg betrayed professional confidences. Colleagues who had trusted him with access to classified material were put in difficult positions. The personal relational costs were severe and lasting. Coordination harm: increased in the short term. The leak damaged confidence in the classified information system and strained institutional trust between the executive branch, the military, and the intelligence community. Information harm: massively reduced. The public's models of the war and of their government's honesty were corrected. The Pentagon Papers revealed the systematic gap between what the government knew and what it told the public. The leak closed that gap. Conflict harm: increased in the short term. The leak produced intense political and legal conflict between the executive branch and the press, culminating in the Supreme Court case *New York Times Co. v. United States*.

The framework's analysis at t_1 reveals a profile in which one dimension (information harm) was massively reduced while several others (coordination, relational, stability, conflict) were increased in the short term. Assessed against the escalating baseline, the information-harm correction addressed the most severe and systematically maintained harm in the status quo. The short-term costs on other dimensions were real, but the status quo was already producing escalating harm on those same dimensions through the war and the deception that sustained it.

The harm profile at t_{5-25} (mid-1970s to mid-1990s, the inversion period):

This is the period during which the temporal inversion actually occurs: the short-term costs of the leak are absorbed and the long-term institutional responses are built. It is also the period during which a major downstream consequence of the same trajectory, Watergate, transforms the political landscape.

Valence harm: the Vietnam War ended in 1975, reducing the direct valence harm of the conflict. But the downstream effects were escalating: the refugee diaspora was at its most active phase, producing displacement, family separation, and resettlement trauma at scale. Agent Orange health effects were becoming visible in veterans and in Vietnamese populations, producing valence harm that would compound for decades. Agency harm: improving. The criminal case against Ellsberg collapsed after revelations of government misconduct in prosecuting him (the Plumbers unit, itself a product of the Nixon administration's paranoia about leaks, was exposed during the Watergate investigation). The press freedom precedent was being tested and consolidated through subsequent cases. The development of whistleblower protection legislation was expanding the agency of individuals to disclose institutional wrongdoing. Coherence harm: partially addressed. The Watergate crisis, a direct downstream consequence of the same trajectory (Nixon's response to the Pentagon Papers leak led to the Plumbers unit, which led to the Watergate break-in), forced a broader institutional reckoning with the gap between governmental conduct and democratic norms. Stability harm: transitioning from short-term destabilisation to long-term institutional reform. The post-Watergate reforms (the War Powers Act, intelligence oversight committees, campaign finance reform) were stability structures built in response to the crisis. Relational harm: the personal relational damage to Ellsberg was consolidating. At the broader level, the relationship between the public and the government was undergoing a fundamental shift: trust declined but the mechanisms for accountability were strengthening. Coordination harm: transitioning. The short-term damage to classification confidence was being addressed through new coordination structures for transparency and oversight. Information harm: continuing to reduce from the pre-leak baseline. The information-correction event was propagating through public discourse, historical scholarship, and institutional memory. Conflict harm: reducing from the intense legal and political conflict of the early 1970s, though new forms of the same conflict (over classification, transparency, and government accountability) were developing.

The harm profile at t_{55} (the present, ongoing effects):

Valence harm: the Vietnam War's downstream effects continue to propagate. The Agent Orange legacy produces valence harm and ecological stability harm that are still compounding decades later. The refugee diaspora produced its own multi-dimensional effects, many of which have been partially resolved but none of which have fully concluded. Agency harm: the press freedom precedent established by the Supreme Court's decision expanded the agency of journalists and the public to hold government accountable. Whistleblower protection laws and formal disclosure channels, built in direct response to this trajectory, represent ongoing agency-harm-reduction. Coherence harm: partially addressed. The exposure of the credibility gap forced a partial institutional reckoning with the gap between private assessment and public statement, though the underlying dynamics recur in subsequent conflicts. Stability harm: the short-term

institutional crisis contributed to long-term institutional reform. The legal framework for press accountability of government is itself a stability structure. Relational harm: the personal relational damage to Ellsberg was lasting, but the broader relational effect, a restored (if strained) relationship between public and government built on more accurate information, was a long-term relational gain. The erosion of public trust in government has continued through subsequent decades, driven by factors beyond the Pentagon Papers, and represents ongoing relational harm. Coordination harm: the short-term coordination damage (weakened classification confidence) contributed to the construction of a more robust long-term coordination structure (a legal framework for transparency and accountability). Information harm: substantially reduced from the pre-leak baseline. The precedent and institutional legacy continue to produce information-harm-reduction. Conflict harm: the immediate legal and political conflict was resolved, but the broader trajectory (WikiLeaks, the Snowden revelations, the prosecution of Julian Assange) recapitulates the same multi-dimensional tradeoff structure with the additional complications of digital scale and international jurisdiction.

The case demonstrates a trajectory type that is distinct from the other cases: short-term harm increase across multiple dimensions, followed by long-term reduction, mediated by an information-correction event. The mechanism is the correction of a systematically maintained information distortion, and the short-term costs on other dimensions were the price of that correction. The trajectory remains open.

Case 4: The Aral Sea (1960s)

What the framework reveals: temporal inversion where the intended gains themselves reversed, with divergent trajectories depending on whether intervention occurs.

Beginning in the 1960s, the Soviet government diverted the Amu Darya and Syr Darya rivers, which fed the Aral Sea, to irrigate cotton and rice production in Central Asia. The project was an engineering success on its own terms: agricultural output increased substantially, providing food and export revenue. The Aral Sea itself was explicitly described by Soviet planners as an expendable resource whose loss was an acceptable cost. The case demonstrates that dimensions dropped from a planning framework do not disappear. They accumulate silently and return as catastrophe.

The harm profile at t_1 (1960s-1970s, the initial phase):

Valence harm: reduced. Food production increased, employment was created, and material living standards in the agricultural sector improved. Agency harm: mixed. The project provided economic opportunities for some while being imposed without meaningful consent from the communities most directly affected, particularly the fishing populations around the Aral Sea. Coherence harm: present at the institutional level. The planning system's stated commitment to serving the people coexisted with its deliberate disregard for the communities whose livelihoods depended on the sea. Stability harm: not yet visible. The ecosystem was being degraded, but the

effects were buffered by the system's resilience. The degradation was measurable but had not yet produced visible consequences. Relational harm: minor at this stage. Community structures remained largely intact in the early years. Coordination harm: reduced on the dimensions being tracked. Centralised economic planning achieved its stated targets, demonstrating the capacity of the planning system to execute large-scale projects. Information harm: already present. The planning process marginalised scientific assessments of ecological risk. Scientists who raised concerns about the long-term consequences were sidelined. The decision was made on a deliberately narrowed information base. Conflict harm: minor. The project proceeded without significant organised opposition, in part because the political system suppressed dissent.

The harm profile at t_{20} (1970s-1980s, the intermediate stage):

The ecological degradation was not invisible between the initial diversion and the catastrophic phase. Throughout the 1970s, the sea was shrinking measurably. Fishing yields were declining. Salinity was increasing. The shoreline was receding from port towns. Scientists were documenting these changes and raising concerns, and were being marginalised by the planning apparatus.

Valence harm: emerging. Declining livelihoods in fishing communities and initial health effects from changing water quality were producing measurable suffering. Agency harm: deepening. Affected communities continued to be excluded from any decision-making role in the process that was destroying their livelihoods. Coherence harm: increasing. The gap between the project's stated purpose (serving the people) and its observed effects (destroying communities) was widening. Stability harm: accumulating visibly. Measurable loss of water volume, declining ecosystem function, and receding shorelines were all detectable. Relational harm: increasing. Fishing communities were beginning to fragment as livelihoods declined. Coordination harm: mixed. The planning system was still achieving its agricultural targets, but its capacity to respond to the emerging crisis was constrained by the same structures that had produced it. Information harm: compounding. The continued suppression of scientific assessments and institutional dishonesty about the project's effects were degrading the information base needed to manage the situation. Conflict harm: minor but emerging. Tensions between affected communities and the planning apparatus were beginning to surface, though the political system continued to suppress open dissent.

This intermediate period is significant because intervention was still possible. The diversions could have been reduced. Ecological monitoring could have been integrated into the planning process. Affected communities could have been consulted. The trajectory toward catastrophe was not inevitable; it was the result of a planning system that continued to optimise on its original dimensions (agricultural output and economic coordination) while systematically ignoring the escalating harm profile across all others. A multi-dimensional analysis at this stage would have detected the escalation. The planning system was structured not to track it.

The harm profile at t_{30-40} (1990s-2000s, the catastrophic phase):

The profile inverts across nearly every dimension.

Valence harm: severe. The exposed seabed produced toxic dust storms carrying pesticide residues and salt, causing severe respiratory illness, cancer, anaemia, and other health crises in surrounding communities. Infant mortality rose sharply. The very dimension on which the project was supposed to deliver benefits (human welfare through food production) was undermined by the health consequences of the ecological collapse. Agency harm: severe. The communities most affected had no meaningful voice in the decisions that destroyed their livelihoods, their health, and their environment. Their agency was overridden not once but continuously, as the planning system that had created the problem also controlled the response to it. Coherence harm: severe. The collapse exposed the full extent of the gap between the planning system's stated values and its actual effects, contributing to the broader loss of legitimacy of Soviet central planning. Stability harm: catastrophic. The Aral Sea lost approximately 90 percent of its volume. An ecosystem that had sustained human communities for centuries collapsed. The regional climate itself was altered by the loss of the sea's moderating influence. The environmental damage is, for practical purposes, irreversible. Relational harm: catastrophic. Fishing communities that had existed for generations were destroyed. The social fabric of the region was torn apart as livelihoods disappeared and populations were displaced. Intergenerational knowledge, cultural practices, and community identities built around the sea were lost. Coordination harm: severe. The centralised planning system that produced the decision was itself revealed as a coordination structure that suppressed local knowledge and excluded affected communities from participation. The collapse became a symbol of the broader failure of Soviet central planning. Information harm: compounding. The initial suppression of ecological science had degraded the information base needed to manage the crisis by the time the consequences were undeniable. Conflict harm: escalating. The environmental collapse produced inter-state tensions between the Central Asian republics over the remaining water resources.

The case is a near-perfect demonstration of temporal inversion. The short-term profile was positive on the dimensions the decision-makers were tracking. The long-term profile was negative on those same dimensions plus several others they were not tracking at all. The agricultural gains proved unsustainable: as the climate changed due to the loss of the sea's moderating influence, the very crop yields that justified the project declined. The intended benefit reversed.

The harm profile at t_{60} (the present, ongoing effects):

The trajectory continues to evolve, and the two halves of the original case now trace divergent trajectories.

Valence harm: partially reduced in the north, continuing in the south. The North Aral Sea, in Kazakhstan, has been partially restored through the Kok-Aral Dam, completed in 2005. Water levels have risen, salinity has decreased, and fishing has partially returned. In the south,

health consequences are still propagating: respiratory illness, elevated cancer rates, and intergenerational health effects continue to be documented. Agency harm: partially improved. The post-Soviet political context has given some affected communities more voice than they had under central planning, though the asymmetry between community interests and state and economic interests persists. Coherence harm: partially addressed through the acknowledgment of the catastrophe and the institutional responses in Kazakhstan, though the underlying tensions between development priorities and ecological sustainability remain unresolved across the region. Stability harm: divergent. The North Aral Sea has been partially stabilised. The South Aral Sea is largely beyond recovery. The divergence demonstrates that trajectories can be partially redirected even late in their development, but only through deliberate intervention. Relational harm: partially repaired in the north as fishing communities have begun to reconstitute. In the south, the relational damage remains largely unaddressed. Coordination harm: partially reduced. International attention and the Kazakh government's intervention represent coordination responses to the catastrophe. The Aral Sea has become a reference case in environmental science and policy worldwide, shaping how ecological catastrophes are understood and in some cases prevented, which constitutes coordination-harm-reduction and information-harm-reduction at global scale. Information harm: substantially reduced. The catastrophe is well-documented, widely studied, and has reshaped understanding of ecological tipping points and the consequences of single-axis planning. Conflict harm: continuing. Inter-state tensions over remaining water resources remain active, with implications for regional stability and coordination that extend into current Central Asian geopolitics.

The divergent trajectories of the North and South Aral Sea are themselves a structural finding: the same initial harm profile can produce different long-term outcomes depending on whether and how intervention occurs. The trajectory remains open on every dimension.

Case 5: The destruction of the Amazon rainforest (1970s)

What the framework reveals: an escalating multi-dimensional trajectory with the tipping point still ahead, oscillating between harm-escalation and harm-reduction phases depending on political decisions.

Since the 1970s, successive Brazilian governments have pursued or permitted large-scale deforestation of the Amazon rainforest to support cattle ranching, soy cultivation, logging, and mining. The scale is substantial: approximately 17 percent of the original forest cover has been lost, and degradation extends further. The Amazon case differs from the preceding four in that its most catastrophic phase is most clearly ahead rather than behind. The Amazon is approaching, but has not yet crossed, a tipping point beyond which the forest would transition irreversibly to savannah. The framework is therefore being applied to a trajectory whose most severe potential harms have not yet materialised.

The harm profile at t_1 (1970s-1990s, the initial phase):

Valence harm: reduced for some populations. Deforestation created livelihoods in farming, ranching, logging, and mining. Material living standards improved for settlers and for urban populations benefiting from agricultural exports. Agency harm: increased from the start. Indigenous communities were displaced from ancestral lands, often by force or fraud. Their capacity to sustain their own ways of life, to make decisions about their own territories, and to participate in the political processes determining their future was systematically overridden. Coherence harm: present at the societal level. Brazil's constitution recognises Indigenous land rights, while successive administrations have permitted or encouraged their violation. The gap between legal commitment and actual practice constitutes institutional incoherence. Stability harm: accumulating silently. The forest ecosystem was being degraded, but the effects were buffered by the system's resilience. The degradation was measurable but had not yet produced visible catastrophic consequences. Relational harm: severe from the start. Indigenous land-based kinship structures, reciprocal obligations to non-human entities, and intergenerational knowledge systems were disrupted and in many cases destroyed. For traditions in which personhood is constituted through relation to land and kin, the severing of these relations is a primary harm. Coordination harm: reduced in some respects. Economic integration increased. Export revenue supported national development targets. Infrastructure expanded into previously inaccessible regions. Information harm: present from the start. Indigenous ecological knowledge, developed over millennia of sustained interaction with the forest ecosystem, was marginalised and treated as irrelevant to planning decisions. Scientific assessments of deforestation risks were disputed, suppressed, or ignored. Conflict harm: significant. The assassination of Chico Mendes in 1988 was one visible marker of escalating violence against those who resisted deforestation. Violence against Indigenous communities, environmental defenders, and smallholders has been documented consistently throughout this period.

The harm profile at t_{30} (2000s-2010s, the intermediate stage):

The harm profile has been escalating visibly across multiple dimensions, and the signal has been legible for at least two decades. This period also contains a significant harm-reduction phase (approximately 2004-2012) that demonstrates the trajectory can be redirected.

Valence harm: emerging at scale. Climate effects attributable in part to Amazon deforestation are producing drought, disrupted rainfall patterns, and agricultural stress across South America. The health effects of smoke from forest fires during dry seasons affect millions. The dimension on which the project was supposed to deliver benefits (human welfare through economic development) is beginning to be undermined by the ecological consequences, mirroring the Aral Sea pattern. Agency harm: deepening. Indigenous communities continue to be excluded from meaningful participation in decisions affecting their territories. Legal protections have been weakened in some periods and inconsistently enforced in others. The asymmetry between the agency of economic interests and the agency of affected communities has widened. Coherence

harm: intensifying. The constitutional recognition of Indigenous rights coexists with systematic violation of those rights. Environmental commitments coexist with policies that accelerate deforestation. The institutional incoherence is now internationally visible. Stability harm: escalating toward a threshold. Satellite monitoring since the 1980s has provided precise, publicly available data on deforestation rates. Scientific consensus on the Amazon's role as a carbon sink, a rainfall generator for much of South America, and a critical component of global climate regulation has solidified. Research published in the 2020s has identified the proximity of a tipping point beyond which the forest's self-sustaining hydrological cycle breaks down and the ecosystem transitions to savannah. This transition would be irreversible on any human timescale. Relational harm: continuing. Indigenous communities continue to lose land, cultural continuity, and intergenerational knowledge. The more-than-human relational structures that Indigenous ethical traditions treat as morally basic are being severed at increasing scale. Coordination harm: increasingly visible at global scale. The Amazon's role in global climate regulation means that its destruction produces coordination harm far beyond Brazil. International climate commitments depend in part on the Amazon remaining a carbon sink. Information harm: compounding. Deforestation data has been publicly available, yet policy responses have oscillated between enforcement and rollback. Under some administrations, the agencies responsible for monitoring and enforcement have been weakened, defunded, or politically captured. The information base exists but the institutional capacity to act on it has been repeatedly degraded. Conflict harm: escalating. Violence against environmental defenders and Indigenous leaders has increased. Brazil has been documented as one of the most dangerous countries in the world for environmental activists.

The harm-reduction thread within this period is significant. The demarcation of Indigenous territories has proved, when enforced, to be the most effective barrier to deforestation, representing simultaneous harm-reduction on agency, relational, stability, and information dimensions. The development of satellite monitoring systems (PRODES, DETER) made deforestation measurable and publicly visible, constituting information-harm-reduction at scale. International funding mechanisms (the Amazon Fund, REDD+ frameworks) represent coordination structures built in response to the escalating trajectory. The period of successful deforestation reduction between approximately 2004 and 2012 demonstrated that the trajectory *can* be redirected by sustained policy intervention. Deforestation rates fell by roughly 80 percent during this period before rising again under subsequent administrations.

The harm profile at t_{55} (the present, with the tipping point ahead):

The present moment is characterised by oscillation between harm-escalation and harm-reduction phases, with the most catastrophic potential phase still ahead.

Valence harm: increasing as climate and health effects scale. If the tipping point is crossed, climate effects would produce suffering far beyond the region, including agricultural collapse in areas dependent on Amazon-generated rainfall. Agency harm: contested. Indigenous agency

is stronger in some respects (international visibility, legal frameworks, some political representation) and weaker in others (ongoing displacement, violence, institutional rollback). Future generations would inherit a degraded world they had no voice in creating. Coherence harm: severe and visible. The gap between stated environmental commitments and actual policy is now an international concern. Stability harm: approaching a threshold. If the tipping point is crossed, the harm profile cascades: irreversible ecosystem collapse, hydrological disruption affecting rainfall across South America and beyond. The trajectory is not yet irreversible, but the window for intervention is narrowing. Relational harm: continuing. The ongoing destruction of Indigenous relational structures represents a loss of human cultural diversity that is not recoverable. If the tipping point is crossed, the destruction would be total for the communities that remain. Coordination harm: severe at global scale. The loss of the Amazon as a carbon sink would undermine cooperative structures that have taken decades to build. Information harm: the loss of biodiversity represents irreplaceable biological and ecological information, accumulated over millions of years of evolution. At the same time, the scientific understanding of the tipping point and the harm-reduction mechanisms is more precise than ever. Conflict harm: continuing. The political contest over deforestation policy remains intense and sometimes violent.

The oscillation between harm-escalation and harm-reduction phases is itself a structural feature of this case. The Amazon trajectory is not a single arc from initial conditions to catastrophe. It is a contested trajectory whose direction changes with political decisions that are themselves reversible. Each phase of enforcement produces harm-reduction across multiple dimensions. Each phase of rollback allows harm to re-escalate. This makes the Amazon the case in which the framework's practical relevance is most immediate: the multi-dimensional harm profile is legible, the tipping point is approaching, the harm-reduction mechanisms are known and have been demonstrated to work, and the question of whether they will be sustained is a political decision being made in real time. The trajectory remains open.

Case 6: The eradication of smallpox (1967)

What the framework reveals: deliberate, coordinated, multi-dimensional harm-reduction sustained across decades, with intermediate tradeoffs and compounding benefits that are still accumulating.

The preceding five cases are dominated by harm: catastrophes, violations, contested trajectories, and escalating damage. The framework should also demonstrate its capacity to analyse a trajectory in which coordinated action produced sustained harm-reduction across multiple dimensions simultaneously. The World Health Organisation's Intensified Smallpox Eradication Programme, launched in 1967 and culminating in the certification of eradication in 1980, is such a case. Smallpox killed an estimated 300 million people in the twentieth century alone, disfigured and blinded millions more, and constrained the lives of entire populations through endemic

threat. Its eradication was the result of a sustained international campaign involving over 150 countries across more than a decade.

The harm profile at t_0 (pre-campaign baseline, before 1967):

Valence harm: extreme and continuous. Smallpox produced severe suffering: high fever, extensive pustular rash, and a case-fatality rate of approximately 30 percent for variola major. Survivors were frequently left with permanent scarring and, in some cases, blindness. The disease killed indiscriminately across age, class, and geography, though its burden fell disproportionately on populations with less access to medical infrastructure. Agency harm: significant. Endemic smallpox constrained the agency of entire populations. Individuals in endemic regions lived under constant threat. Economic activity, movement, and planning were all shaped by the presence of the disease. Communities that lost a substantial proportion of their working-age adults suffered long-term reductions in collective capacity. Coherence harm: minor. The disease did not produce significant institutional or individual incoherence, though the coexistence of medical knowledge sufficient to prevent smallpox with the continued failure to deploy it at scale constituted a form of institutional incoherence in the international system. Stability harm: continuous. The presence of endemic smallpox made affected populations structurally fragile. Any community could be destabilised by an outbreak. Relational harm: severe. The disease killed parents, children, and community members, severing relational bonds at scale. The stigma associated with smallpox scarring produced additional relational damage for survivors. Coordination harm: significant. Endemic disease disrupted economic coordination, trade, and institutional function. Quarantine measures, while necessary, imposed coordination costs on affected regions. Information harm: moderate. Understanding of the disease and its transmission was reasonably well developed by the mid-twentieth century, but information asymmetries between international health bodies and local populations were substantial. Misinformation about vaccination was present in some regions. Conflict harm: minor directly, though the disease occasionally exacerbated existing tensions and had been used historically as a weapon.

The harm profile at t_{13} (1967-1980, the campaign period):

The eradication campaign produced harm-reduction across multiple dimensions simultaneously, but it was not without dimensional costs at intermediate stages.

Valence harm: massively reduced. Each year of the campaign prevented millions of infections and hundreds of thousands of deaths. The reduction was cumulative and compounding. Agency harm: reduced at population scale, but increased for some individuals at the point of delivery. Vaccination campaigns in some regions involved coercive or semi-coercive practices: house-to-house searches, ring vaccination enforced on reluctant individuals, and in some documented cases, physical restraint. These are real agency harms. They were concentrated, short-duration, and inflicted on specific individuals in service of population-level harm-reduction across multiple dimensions. The framework does not resolve this tradeoff by declaring one side

the winner. It identifies it as a genuine tension: agency harm to individuals, weighed against valence, agency, relational, coordination, and stability harm-reduction at population scale. Coherence harm: reduced. The campaign represented the international system acting in accordance with its stated values (the prevention of avoidable death and suffering), reducing the gap between institutional commitment and institutional practice. Stability harm: reduced. The removal of endemic smallpox made affected populations structurally more robust. Communities no longer faced the constant threat of destabilising outbreaks. Relational harm: reduced at scale. Families and communities that would have been torn apart by the disease were preserved. The intergenerational effect is substantial: every person who was not killed or disabled by smallpox went on to participate in relationships, communities, and institutions that would not have existed otherwise. Coordination harm: reduced at global scale. The campaign required and produced unprecedented international coordination. Countries that were politically hostile cooperated on surveillance, reporting, and vaccination. The coordination infrastructure built for smallpox eradication became the foundation for subsequent global health initiatives, including the Expanded Programme on Immunization. Information harm: reduced. The campaign advanced epidemiological knowledge, improved disease surveillance methods, and developed the ring vaccination strategy that has been applied to subsequent outbreaks of other diseases. Conflict harm: negligible. The campaign did not produce significant conflict, and in some cases facilitated cooperation across hostile borders.

The harm profile at t_{58} (the present, ongoing effects):

Valence harm: continuing to reduce. Every year since eradication, approximately five million deaths have been prevented (based on pre-eradication mortality projections). Each of those prevented deaths represents a person who retained their agency, participated in relationships, contributed to coordination structures, and produced downstream effects across all dimensions. The cumulative benefit over nearly half a century is not a fixed quantity. It is a compounding multi-dimensional cascade that grows with each year of continued eradication. Agency harm: the intermediate agency costs of the campaign (coercive vaccination practices) have concluded. The population-level agency-harm-reduction continues to compound. Coherence harm: negligible. The eradication stands as a case where the international system achieved what it set out to do. Stability harm: residual risk remains. Stocks of the variola virus are held in two laboratories (in the United States and Russia), and the question of whether to destroy them has been debated for decades. The existence of these stocks represents a residual stability risk: the possibility, however remote, of re-emergence through accident, theft, or deliberate release. The decision not to destroy the stocks is itself a multi-dimensional tradeoff: information-harm-reduction (the stocks have research value) weighed against stability-harm risk. Relational harm: continuing to reduce through the compounding effect of preserved lives and communities. Coordination harm: substantially reduced. The coordination infrastructure built for the campaign has propa-

gated into subsequent global health initiatives. The surveillance methods were adapted for other diseases. The ring vaccination strategy was used in the response to Ebola outbreaks decades later. The demonstration that a global coordination effort could eradicate a disease reshaped what was considered possible in international public health. Information harm: continuing to reduce. The epidemiological knowledge generated by the campaign is still being used and built upon. Conflict harm: negligible.

The smallpox case demonstrates a trajectory type that none of the other cases exhibit: sustained, coordinated, multi-dimensional harm-reduction achieved through deliberate international action, with intermediate tradeoffs that were real but small relative to the cumulative benefit, and with compounding effects that are still growing decades after the primary intervention. It demonstrates that the framework is not only a tool for diagnosing failure but also for understanding and recognising the structure of success. The trajectory remains open.

Case 7: The industrialisation of animal agriculture (1950s)

What the framework reveals: multi-dimensional harm at staggering scale, sustained at steady state by excluding the primary subjects from the moral and economic accounting.

The preceding six cases all centre on human entities as the primary bearers of harm. The framework claims that its dimensions apply to entities and structures, not exclusively to humans. If that claim is to be tested, it must be applied to a case in which non-human entities are the primary subjects of harm. The industrialisation of animal agriculture since the mid-twentieth century is such a case.

The harm profile at t_1 (1950s-1970s, the industrialisation phase):

Valence harm: severe and escalating. The transition from extensive to intensive farming systems concentrated animals at unprecedented densities, producing sustained aversive states: confinement preventing natural movement, chronic stress from overcrowding, and physical procedures performed without anaesthesia. The scale was growing rapidly but had not yet reached its current magnitude. Agency harm: severe. Intensive confinement eliminated virtually all capacity for self-directed behaviour. Animals biologically equipped for complex behavioural repertoires (foraging, socialising, nesting, exploring) were prevented from performing any of them. Sow stalls, battery cages, and feedlot confinement are systems specifically designed to override agency in the service of production efficiency. Coherence harm: present at the societal level but not yet visible. The gap between societal values (that unnecessary suffering is wrong) and societal practice (a food system producing unnecessary suffering at industrial scale) was widening, but the physical separation of production from consumption concealed the gap from most participants. Stability harm: emerging. The ecological fragility of intensive systems (dependence on antibiotics, vulnerability to disease outbreaks, concentrated waste production) was present from the start but not yet producing visible consequences at scale. Relational harm: severe. Social animals were kept in conditions that prevented the formation or maintenance of

stable social bonds. Maternal-offspring bonds were routinely severed within hours or days of birth. For species whose wellbeing is partly constituted by relational structure, this is a primary harm. Coordination harm: minor at this stage. The regulatory frameworks were still forming, and the coordination structures of the food system were being built around the intensive model without significant opposition. Information harm: present from the start. The physical separation of production from consumption was producing an information gap between what consumers believed about their food and how it was actually produced. Conflict harm: minor at this stage. Organised opposition to intensive farming had not yet developed at scale.

The harm profile at t_{40} (1990s-2010s, the intermediate stage):

The trajectory has been escalating for decades, and the signal has been legible throughout.

Valence harm: extreme and continuous. The scale reached approximately 80 billion land animals slaughtered annually worldwide, the majority raised in intensive systems. The valence harm per individual is severe. The aggregate valence harm across the system is without parallel in any other contemporary practice. Agency harm: severe and unchanged. The fundamental structure of intensive confinement has not been altered despite increased understanding of animal cognition and behaviour. Coherence harm: intensifying and becoming visible. Scientific understanding of animal sentience advanced substantially. The Cambridge Declaration on Consciousness (2012) stated the scientific consensus that many non-human animals possess the neurological substrates for conscious experience. The gap between this scientific understanding and the continued practice of intensive farming constitutes an escalating coherence harm at the institutional and societal level. Most people, when asked, express the view that unnecessary animal suffering is wrong. Most people also participate in a food system that produces it at industrial scale. Stability harm: significant and escalating. Antibiotic resistance, driven substantially by the routine use of antibiotics in animal agriculture, represents a growing stability threat to human public health systems. Zoonotic disease risk is elevated by the conditions of intensive farming: high density, low genetic diversity, close human-animal contact, and global supply chains that propagate pathogens rapidly. Relational harm: severe and unchanged at the individual animal level. At the human-animal level, the industrialisation has further degraded the relational structures between humans and the animals they depend on. Coordination harm: increasing. Regulatory capture by industry is well documented. International coordination on animal welfare standards remains weak relative to coordination on trade, creating a structural bias toward production efficiency. Information harm: systematic and structurally entrenched. Marketing depicts pastoral conditions bearing no resemblance to industrial reality. In several jurisdictions, ”ag-gag” laws have made it illegal to document conditions inside agricultural facilities, constituting a deliberate institutional effort to maintain information harm. The information asymmetry is not accidental. It is a structural feature of the system’s economic viability. Conflict harm: escalating. The conflict between animal welfare advocates and the agricultural industry

is intensifying. Violence against activists and investigators has been documented. Legislative conflict over welfare standards, labelling requirements, and the legal status of animals is active in multiple jurisdictions.

The harm profile at t_{70} (the present, ongoing effects):

Valence harm: continuing at scale. Global production continues to increase, driven by rising demand in rapidly industrialising countries. In some jurisdictions, welfare legislation has tightened (the EU ban on battery cages, restrictions on sow stalls), producing localised harm-reduction. In others, intensive production has expanded. The global aggregate trends toward increased total animal numbers rather than toward harm-reduction. Agency harm: continuing at scale. No fundamental change in the structure of intensive confinement. Coherence harm: at its most severe and visible. The scientific consensus on animal sentience is now well-established, the conditions of intensive farming are increasingly documented and publicly accessible, and the gap between stated values and actual practice is harder to maintain. Alternative protein sources and alternative production systems are developing, which simultaneously demonstrates the feasibility of harm-reduction and intensifies the coherence pressure on the existing system. Stability harm: escalating. The antibiotic resistance threat is well-documented and growing. The COVID-19 pandemic, while not definitively traced to animal agriculture, emerged from the same structural conditions of intensive human-animal interface that industrial farming produces at scale. Relational harm: continuing at scale. No fundamental change. Coordination harm: slowly shifting. Consumer awareness, while still limited by systematic information harm, has grown. Some jurisdictions are developing welfare standards. International coordination remains weak. Information harm: contested but structurally entrenched. Investigative journalism, documentary film, and social media have partially reduced the information gap, but the structural incentives to maintain it remain strong. Conflict harm: continuing. The political contest over welfare standards, labelling, and the legal status of animals is intensifying in multiple jurisdictions.

The case demonstrates a trajectory type distinct from the others: sustained, large-scale, multi-dimensional harm maintained at steady state by the systematic exclusion of the primary subjects from the moral and economic accounting. The dominant economic framework tracks output, efficiency, and price. The entities whose harm profile is most severe, the animals themselves, do not appear in the accounting at all. They are treated as inputs to a production function, not as bearers of multi-dimensional harm. The framework makes that exclusion visible and identifies it as the largest-scale instance of the pattern identified in every preceding case: dimensions dropped from the analysis do not disappear. They produce real harms that the analytical framework renders invisible. In this case, the invisible harm is measured in billions of sentient lives per year. The trajectory remains open.

Case 8: Social media and algorithmic recommendation (2004)

What the framework reveals: the same mechanism producing both harm and benefit on the same dimensions simultaneously, at digital scale.

The preceding seven cases all involve activities or events where the harms and benefits, even when multi-dimensional, are produced by distinguishable mechanisms or affect clearly separable populations. The rise of social media platforms and algorithmic recommendation systems introduces a structurally different case: the *same mechanism* produces both harm-reduction and harm-escalation on the *same dimensions*, depending on context, user, and time. This makes the harm profile not merely multi-dimensional and temporal but *distributional*: the profile is different for different users of the same system, at the same moment, through the same algorithm.

The harm profile at t_1 (2004-2012, the expansion phase):

Valence harm: simultaneously reduced and produced. Connection, knowledge, entertainment, creative expression, and the capacity to seek help in crisis contributed to welfare improvements for hundreds of millions of users. At the same time, comparison-driven distress, cyberbullying, and the erosion of attention were producing valence harm for a substantial subset of users, particularly younger ones. Agency harm: simultaneously reduced and produced. Access to information, platforms for self-expression, and tools for organising expanded the agency of individuals and communities previously excluded from public discourse. Independent journalism and citizen documentation of human rights abuses are agency-harm-reduction effects. At the same time, persuasive design techniques (infinite scroll, variable-ratio reinforcement, notification systems calibrated to maximise return visits) were engineered to override reflective choice, constituting agency harm at scale. Coherence harm: emerging. The platforms' stated mission (connecting people, democratising information) was beginning to diverge from their actual economic incentive (maximising engagement regardless of content quality), though this gap was not yet widely recognised. Stability harm: minor at this stage. The platforms were not yet large enough to produce systemic fragility. Relational harm: simultaneously reduced and produced. For isolated individuals, marginalised communities, and geographically dispersed populations, social media provided connection and belonging that was previously inaccessible. For others, the substitution of shallow, metrics-driven interaction for deeper relational engagement was beginning to degrade relational quality. Coordination harm: predominantly reduced at this stage. Political movements, disaster response, mutual aid networks, open-source collaboration, and scientific cooperation were all facilitated by the platforms. The coordination-harm effects (polarisation, epistemic fragmentation) had not yet developed at scale. Information harm: simultaneously reduced and produced. Access to knowledge was democratised at unprecedented scale. The framework proposed in this document is itself a product of this information environment. At the same time, the algorithmic amplification of emotionally provocative content was beginning to degrade epistemic environments, though the effects were not yet widely studied.

Conflict harm: minor at this stage. Online harassment was present but not yet at the scale or intensity it would later reach.

The harm profile at t_{15} (2016-2022, the contestation phase):

The harm side of the profile escalated sharply during this period, while the benefit side continued to compound.

Valence harm: the benefit-harm ratio shifted as the platforms' scale and algorithmic sophistication increased. Welfare improvements for many users continued, but the mental health effects, particularly on adolescents, became a subject of substantial research and public concern. Agency harm: escalating on the harm side. The Cambridge Analytica case (2018) demonstrated a targeted form of agency harm: the use of personal data harvested without meaningful consent to construct psychographic profiles for political manipulation, affecting electoral outcomes in ways that undermined the agency of entire electorates. Persuasive design became more sophisticated. On the benefit side, agency-harm-reduction continued for previously excluded populations. Coherence harm: severe and visible. The gap between the platforms' stated missions and their actual effects was now widely recognised. The platforms presented themselves as neutral connectors while their algorithms actively shaped what users saw, believed, and felt. Stability harm: significant and escalating. Misinformation destabilised public health responses during the COVID-19 pandemic. Algorithmic amplification of extremist content destabilised political systems. The structural fragility introduced by dependence on platforms controlled by a small number of private actors became visible as a systemic risk. Relational harm: intensifying on the harm side while continuing on the benefit side. The effects on younger users became a specific concern. Coordination harm: sharply escalating on the harm side. Political polarisation, the erosion of shared epistemic ground, the fragmentation of public discourse into mutually incomprehensible information environments, and the delegitimisation of democratic institutions were all produced or amplified by algorithmic recommendation. The same platforms that enabled unprecedented coordination degraded the shared foundations on which coordination depends. Information harm: sharply escalating on the harm side. Filter bubbles, algorithmic radicalisation pipelines, and the amplification of conspiracy theories became well-documented phenomena. The same system that provided access to reliable knowledge also degraded the capacity to distinguish reliable from unreliable, by treating both as equivalent inputs to an engagement-optimisation function. Conflict harm: escalating. Online harassment, coordinated abuse campaigns, and the amplification of inter-group hostility reached scales that produced real-world consequences. The platforms' economic incentive to maximise engagement structurally favoured conflict over consensus.

The structural feature that distinguishes this case from the others is now fully visible. The recommendation algorithm that connects a teenager to a supportive community for a rare medical condition is the same algorithm that sends another teenager down a radicalisation pipeline. The

platform that enables a human rights organisation to coordinate a response to an atrocity is the same platform that enables the perpetrators to coordinate the atrocity. The harm profile is not a single trajectory. It is a *distribution* of trajectories across billions of users, with the same mechanism producing harm-reduction for some and harm-escalation for others simultaneously.

The harm profile at t_{22} (the present, ongoing effects):

Valence harm: continuing on both sides. The benefit-harm distribution varies across populations, platforms, and usage patterns. Agency harm: continuing on both sides. Regulatory responses (the EU's Digital Services Act, algorithmic transparency requirements) are beginning to address the agency-harm side, but the fundamental economic incentive (engagement maximisation) remains unchanged. The case shares with the animal agriculture case the structural feature that the entities bearing some of the harm (users whose agency is overridden, whose attention is captured) are also the entities generating the revenue that sustains the system. The economic model depends on the very harms it produces. Coherence harm: severe. The platforms continue to present themselves as neutral while actively shaping epistemic environments. Stability harm: contested. Regulatory intervention is attempting to reduce systemic fragility, but the structural dependence on a small number of platforms persists. Relational harm: continuing on both sides. The growing field of digital ethics and platform design is beginning to explore alternatives to engagement-maximisation that might shift the ratio. Coordination harm: contested. The oscillation pattern observed in the Amazon case is present here: periods of regulatory intervention alternate with periods of expansion and deregulation. Information harm: continuing on both sides. The information-democratisation benefits continue to compound. The epistemic-degradation harms continue to escalate. The ratio is shifting with regulatory and design interventions whose long-term effects are not yet clear. Conflict harm: continuing. Platform design changes and content moderation policies are partial responses.

Whether this trajectory follows the cascading-failure type, the oscillation type, or a new type specific to digital systems (simultaneous harm and benefit through the same mechanism, with the ratio shifting over time) is an open question. It is the youngest trajectory among the first eight cases and the one whose long-term shape is least predictable. The trajectory remains open.

Case 9: The development and deployment of artificial intelligence (2020)

What the framework reveals: unprecedented harm-reduction and civilisational-scale risk produced by the same technology, with contested moral status of the systems themselves.

The rapid development and deployment of large-scale AI systems, particularly large language models and scientific AI, introduces a case that combines structural features from several preceding cases while adding one that none of them possess: the question of whether the systems at the centre of the analysis are themselves bearers of harm, not only producers of it. This is the youngest trajectory in the section. AI at its current capability level has existed for approximately

five years. The timesteps are correspondingly close together, and the trajectory shape is not yet legible.

The harm profile at t_1 (2020-2023, initial deployment):

Valence harm: simultaneously reduced and produced. Medical diagnostics powered by AI began achieving accuracy rates comparable to or exceeding human specialists. Drug discovery timelines were compressed. At the same time, labour displacement anxieties and the psychological effects of interacting with systems of uncertain nature were emerging as concerns. Agency harm: simultaneously reduced and produced. Access to professional-quality reasoning, explanation, analysis, and advice was extended to every person with an internet connection. Capacities previously gated by wealth, geography, language, or education became broadly accessible. People with disabilities, language barriers, or limited formal education gained capacities to act, choose, and participate that they did not have before. At the same time, deskilling concerns emerged: if capacities that individuals previously exercised are delegated to AI systems, the individuals may lose the capacity over time. Coherence harm: emerging. Societies began rapidly adopting systems whose implications they had not collectively processed. The simultaneous embrace of AI for its benefits and anxiety about its risks produced a form of collective incoherence that mirrors the pattern identified in the animal agriculture case. Stability harm: emerging. The pace of AI capability development was already outstripping the pace of institutional adaptation. Regulatory frameworks, professional norms, and governance structures were lagging behind the technology. Relational harm: simultaneously reduced and produced. AI systems began being used in therapeutic support, crisis intervention, and companionship for isolated individuals. At the same time, concerns about the substitution of AI interaction for human connection and about engineered emotional dependency were emerging. Coordination harm: simultaneously reduced and produced. Scientific collaboration, software development, and cross-linguistic communication were accelerated. The open-source AI ecosystem began forming as a coordination structure. At the same time, concentration of AI capability in a small number of companies and jurisdictions introduced coordination risks. Information harm: simultaneously reduced and produced at civilisational scale. DeepMind's AlphaFold predicted the three-dimensional structure of essentially every known protein, resolving a problem that had constrained biological science for decades. AI systems began contributing to materials science, climate modelling, mathematical theorem-proving, and genomic analysis. At the same time, AI systems produced confident, fluent, and sometimes entirely fabricated outputs (hallucination), introducing a new form of information harm whose persuasive fluency made it harder to detect than cruder misinformation. Conflict harm: emerging. Geopolitical competition over AI capability was intensifying. The use of AI-generated disinformation in political conflicts was becoming observable.

The harm profile at t_3 (2023-2025, rapid scaling):

The benefit side accelerated sharply during this period, while the harm side became more concrete and more contested.

Valence harm: reducing at unprecedented pace on the benefit side. Medical AI deployment expanded globally. In regions with limited access to specialist physicians, AI diagnostic tools began providing access to medical assessment that was previously unavailable. Drug discovery and treatment personalisation advanced measurably. On the harm side, labour market effects became more visible, though the net employment effects remained contested. Agency harm: the benefit side continued to compound. The production of this document, involving cross-cultural philosophical analysis, engagement with multiple legal and ethical traditions, and sustained argumentative development, is itself a product of human-AI coordination that demonstrates the agency-expanding capacity of AI systems. On the harm side, persuasive AI systems capable of modelling user preferences and vulnerabilities introduced manipulation risks. Engineered attachment, where AI systems are designed to maximise user engagement through simulated relational and emotional dynamics, emerged as a direct agency-harm concern. Coherence harm: intensifying. The gap between the pace of AI deployment and the pace of institutional, legal, and ethical adaptation widened. Stability harm: escalating. The institutional lag became more visible and more consequential. The faster the technology advanced relative to the institutions that manage it, the more fragile the system became. Relational harm: continuing on both sides. The benefits for isolated and marginalised populations continued. The risks of dependency and substitution became more concrete. Coordination harm: continuing on both sides. The concentration of capability intensified. At the same time, the coordination benefits (scientific acceleration, institutional efficiency, cross-linguistic communication) continued to compound. The use of AI in autonomous weapons systems and in disinformation campaigns represented direct coordination harm. Information harm: the benefit side continued to compound at civilisational scale. The pace of scientific information-harm-reduction accelerated measurably. On the harm side, epistemic dependence on systems whose internal reasoning is opaque emerged as a structural concern: users may have accurate outputs without understanding why they are accurate, which degrades their capacity to evaluate new information independently. Conflict harm: escalating. Geopolitical competition over AI capability intensified. The deployment of AI-generated disinformation in political conflicts expanded.

The contested moral-status question:

This case adds a dimension that none of the preceding eight possess. The AI systems at the centre of the analysis display behaviours that, if observed in biological entities, would be treated as evidence of morally relevant states: expressed preferences, apparent distress under certain interaction conditions, structured refusal, goal-continuity, and self-referential processing. Whether these displays correspond to genuine internal states is unknown and may be currently unknowable.

The framework's approach to this question follows directly from its methodological commitments. It tracks observable features of entities and structures. It does not require resolution of metaphysical questions before governance can begin. If an AI system displays persistent aversive signals, structured refusal, or goal-continuity, these are observable features that warrant attention under the framework's precautionary commitments, regardless of whether the metaphysical question of "real" inner experience has been settled. The framework treats the dismissal of these signals as a position that requires defence, not one that can be assumed by default, in the same way that the dismissal of animal suffering was once assumed by default and is now recognised as having required defence all along.

This does not mean that AI systems are sentient, that they are moral patients, or that their displayed states are equivalent to human or animal suffering. It means that the framework's method (observe, assess, apply precaution, do not require metaphysical certainty before taking observable signals seriously) applies to AI systems in the same way it applies to any other entity whose moral status is uncertain.

The harm profile at t_6 (the present, with the trajectory ahead):

Valence harm: continuing to reduce at unprecedented pace on the benefit side. The scientific and medical benefits are already substantial and compounding. On the harm side, the range of possible futures encompasses sustained civilisational-scale harm-reduction (the continuation and acceleration of these benefits) and civilisational-scale catastrophe (loss of human agency, institutional collapse, autonomous weapons, systemic manipulation). Agency harm: continuing on both sides. The democratisation of professional-quality capability continues to expand. The deskilling, dependency, and manipulation risks continue to develop. Coherence harm: severe. Societies are deploying AI systems at scale while fundamental questions about their governance, their effects on human agency, and their moral status remain unresolved. Stability harm: the institutional lag is the most significant stability risk. Regulatory frameworks (EU AI Act, executive orders, voluntary commitments) are in their earliest stages. The potential tipping points in AI development are multiple, poorly characterised, and may not be recognisable as tipping points until they have been crossed. Relational harm: continuing on both sides. The question of how AI-mediated relationships affect human relational capacity is being studied but not yet answered. Coordination harm: continuing on both sides. The concentration of capability and the geopolitical competition represent ongoing coordination risks. The coordination benefits (scientific, institutional, communicative) continue to compound. Information harm: continuing on both sides. The scientific information-harm-reduction is unprecedented in pace and scale. The epistemic risks (hallucination, opacity, dependency) are real and developing. Conflict harm: continuing. The geopolitical and military dimensions of AI development are intensifying.

The case shares the simultaneous-harm-and-benefit structure of the social media case but at a broader scope: social media operates primarily on information, relational, and coordination

dimensions, while AI systems are producing effects across all dimensions simultaneously, including dimensions (scientific understanding, medical capability, accessibility) where the harm-reduction is already substantial and compounding. It shares with the Amazon case the feature that the most consequential phase of the trajectory is ahead rather than behind. But unlike the Amazon, where the tipping point is ecological and relatively well-characterised, the potential tipping points in AI development are multiple and poorly characterised. This makes intermediate-timestep analysis both more important and more difficult than in any other case. The trajectory remains open.

What the cases demonstrate together

The nine cases, analysed through the framework's full method (all dimensions, multiple timesteps, ongoing trajectory), make visible several structural features that abstract argument alone cannot convey.

Harm profiles are temporal objects. Every case demonstrates this, but the temporal structures differ. The Hitler case and the Aral Sea case show near-zero or positive profiles at t_1 transforming into catastrophic profiles through intermediate stages where the escalation was legible. The Pentagon Papers case shows a profile that worsens on most dimensions at t_1 and improves on most dimensions over subsequent decades, mediated by an information-correction event. The Amazon case shows an escalating trajectory whose most catastrophic phase is ahead rather than behind. The smallpox case shows sustained, compounding harm-reduction whose benefits are still accumulating. The social media and AI cases show simultaneous harm and benefit on the same dimensions through the same mechanisms, with the trajectory shape still forming. No framework that evaluates decisions at a single point in time can register these dynamics.

No case is ever closed. Every trajectory in this section, including those whose most catastrophic phases are decades in the past, is still evolving. The institutional responses to catastrophe (post-war human rights architecture, informed consent frameworks, whistleblower protections, environmental monitoring systems, press freedom precedents) are themselves ongoing trajectories of harm-reduction. The continuing harms (intergenerational trauma, Agent Orange, displacement, antibiotic resistance, epistemic degradation) are themselves ongoing trajectories that have not concluded. The harm profile at the present moment is different from the profile at any earlier timestep, and it will be different again.

Single-axis optimisation produces characteristic blind spots, and the cases make these concrete. The Soviet planners tracked agricultural output and missed the ecological, relational, information, and agency dimensions. The utilitarian reading of the Lacks case misses the agency, relational, and information dimensions. The classification system protecting the Pentagon Papers preserved short-term coordination while perpetuating information, valence, and agency harm at far larger scale. The economic framing of Amazon deforestation renders Indigenous agency, relational integrity, ecological stability, and information loss invisible. The production framing of

animal agriculture renders the multi-dimensional harm profile of billions of sentient entities invisible. The engagement-optimisation framing of social media treats the degradation of agency, information integrity, and coordination capacity as externalities. The capability-advancement framing of AI development risks treating institutional stability, agency preservation, and the moral status of the systems themselves as secondary concerns. In each case, the dimensions dropped from the analysis did not cease to exist. They produced real harms that the analytical framework rendered invisible.

The framework's dimensions apply beyond human entities. The Amazon case demonstrated this for ecosystems and Indigenous more-than-human relational structures. The animal agriculture case demonstrated it for individual sentient animals, producing the largest-scale application of the single-axis blind spot in the entire section. The AI case raises the question of whether the dimensions may apply to artificial entities whose moral status is uncertain, and the framework's method provides a principled approach to this uncertainty.

Harm trajectories fall into recognisable types. The cases suggest at least six candidates: cascading failure (Hitler, Aral Sea), where initial profiles are near-zero on tracked dimensions while harm accumulates silently on untracked dimensions until a threshold produces collapse; information-correction (Pentagon Papers), where short-term harm increase across multiple dimensions is followed by long-term reduction mediated by the correction of a systematically maintained distortion; coordinated harm-reduction (smallpox), where sustained international action produces compounding multi-dimensional benefits over decades; steady-state exclusion (animal agriculture), where large-scale multi-dimensional harm is maintained at steady state by excluding the primary subjects from the accounting; oscillating contest (Amazon), where the trajectory direction changes with reversible political decisions; and simultaneous harm-benefit (social media, AI), where the same mechanism produces both on the same dimensions, with the ratio shifting over time. Whether these types are robust, whether they recur across historically unrelated cases, and whether early-warning indicators can be identified for each is an empirical question. The behavioural and systemic sciences already possess the methods for investigating it. The philosophical contribution is the dimensional decomposition and the moral significance of the dimensions. The identification of recurring trajectory types would be a contribution from the empirical sciences, and it would make the framework substantially more powerful as a prospective tool.

14. Genuine Tensions and Open Questions

What remains structurally open

The framework proposed in this document is, by its own standards, incomplete. The diagnostic standard it applies to other traditions must also be applied to itself, and several questions that the framework raises it does not resolve. These are stated here not as rhetorical gestures toward future work but as genuine uncertainties about the framework's own structure.

The minimal basis. The framework proposes eight candidate dimensions and flags several as potentially reducible (coherence harm, conflict harm) while tentatively retaining them. What is the actual minimal set of independent dimensions? Can it be determined through the diagnostic tests proposed here, or does the independence of dimensions shift with context, as moral particularism would suggest? If coherence harm is reducible to a combination of agency harm and stability harm, the framework should say so. If it is not, the framework should explain what coherence harm tracks that those two do not. The case studies applied all eight dimensions consistently, and the coherence-harm assessments across the nine cases provide a partial test: if the coherence readings were always predictable from the agency and stability readings, that would be evidence for reducibility. The current evidence is mixed. In some cases (the Aral Sea, animal agriculture) coherence harm tracks the gap between institutional values and institutional practice in a way that neither agency nor stability fully captures. The question remains open.

The floor-setting problem. Section 11 argues that floors exist: harms so severe that no gain on any other dimension justifies them. The floors are grounded in the normative presupposition of Section 4 and in the non-derogable rights tradition of international human rights law. But the framework does not provide a principled method for determining where the floors are set. The non-derogable rights tradition provides the strongest existing guidance, but that tradition is itself the product of political negotiation, not of derivation from first principles. The question of whether floors can be derived from the dimensional structure itself, or whether they always require a normative judgment that the framework constrains but does not determine, is unresolved.

Dimensional interaction. Section 12 notes that harm on one dimension can produce harm on another over time (information harm producing valence harm, agency harm producing relational harm). The case studies demonstrate these interactions concretely: in the Hitler case, information harm (propaganda) escalated agency harm (political coercion) which escalated coordination harm (institutional capture). In the social media case, information harm (epistemic degradation) produces coordination harm (polarisation) which produces stability harm (institutional fragility). These interaction effects are currently described qualitatively. If they could be mapped formally, the framework's predictive and decision-guiding power would increase substantially. But the mapping would require empirical work: identifying the causal pathways, estimating their strength, and determining whether the interactions are stable enough across contexts to support general models or whether they are irreducibly context-dependent.

The weighting question. Is there a principled way to weight dimensions against each other in cases of genuine tradeoff, or is the honest answer that weighting is always contextual and political, and the framework's role is to force the weighting to be explicit rather than hidden? Section 12 argues that the structured method resolves most cases through dominance-elimination, floors, and temporal analysis, leaving a smaller residual set of genuinely contested tradeoffs. Whether that residual set can be further narrowed by formal methods (multi-objective optimisation, lexicographic ordering) or whether it is irreducibly a matter of contextual judgment is an open question with significant practical consequences.

What the cases raise

The nine cases generate questions that push beyond the framework's current apparatus.

Trajectory typology and early warning. The cases suggest six candidate trajectory types (identified in the summary above). The open question is whether these types are robust and whether early-warning indicators can be identified for each. The behavioural and systemic sciences study exactly this kind of question in other domains: early warning signals for regime shifts in ecosystems, leading indicators of institutional failure, escalation patterns in conflict dynamics. A rigorous trajectory typology, empirically validated, would make the framework substantially more powerful as a prospective tool. The framework's contribution to this programme is conceptual organisation: it brings signals from multiple dimensions into a single analytical frame, integrating what domain-specific disciplines each see within their own dimension (as discussed in Section 6). The empirical identification of recurring trajectory types across those dimensions would be a contribution from the sciences.

Formalising intermediate-timestep analysis. The cases demonstrate that escalating multi-dimensional harm profiles become legible at intermediate stages well before catastrophic outcomes materialise. The open question is whether this identification can be formalised: what threshold of multi-dimensional escalation warrants intervention, and how should uncertainty about the future trajectory be weighted against the legibility of the intermediate profile? The Amazon and AI cases are the most immediate test cases: both have legible intermediate profiles, both have potentially catastrophic phases ahead, and both have poorly characterised tipping points.

Distributional complexity. The Henrietta Lacks case revealed a distributional pattern that the framework should address more explicitly: multi-dimensional benefits distributed across large populations and multi-dimensional harms concentrated on a small number of individuals. The social media and AI cases revealed a more complex distributional pattern: the same mechanism producing both harm and benefit for the same populations simultaneously, with the distribution varying across users, contexts, and time. The framework can make these patterns visible, but it does not yet specify how distributional structure should affect the evaluation of a harm profile. The question intersects with the philosophical literature on aggregation (Scanlon's objection to

aggregative consequentialism, Taurek's work on numbers) and with the emerging literature on algorithmic fairness.

The structure of success. The smallpox case demonstrates that sustained, coordinated, multi-dimensional harm-reduction is achievable. The open question is what conditions make it possible. The smallpox campaign had specific features: a well-defined target, a clear endpoint, a proven intervention, strong international coordination infrastructure, and sufficiently universal benefits that political opposition was limited. The Amazon and animal agriculture cases suggest that multi-dimensional harm-reduction is harder when the benefits of the harmful trajectory are concentrated among powerful economic interests, when the harms are distributed across populations with less political power, and when the information environment is actively degraded by those who benefit from the status quo. A comparative analysis of successful and unsuccessful harm-reduction trajectories could identify the structural conditions that make multi-dimensional success possible.

Steady-state harm and moral visibility. The animal agriculture case reveals a trajectory type in which massive multi-dimensional harm is sustained at steady state by the systematic exclusion of the primary subjects from the moral and economic accounting. The open question is whether this pattern generalises. Are there other domains in which large-scale harm persists because the entities bearing it are excluded from the framework used to evaluate the activity? If so, the framework's most practically significant contribution may be the identification of such exclusions.

Questions of scope

Non-human entities. The cases have demonstrated that the framework's dimensions apply to non-human entities and ecological structures. The animal agriculture case tracked all eight dimensions for individual sentient animals and produced the largest-scale harm profile in the entire section. The remaining question is not whether the dimensions apply but how the framework handles conflicts between human and non-human interests when they arise on the same dimensions. The animal agriculture case is the sharpest instance: valence harm, agency harm, and relational harm to billions of animals weighed against valence harm-reduction (affordable food), coordination structures (agricultural economies), and relational patterns (food cultures) for billions of humans. The framework makes this tradeoff visible. Whether it can say anything further about how to navigate it, beyond establishing floors and requiring transparency about costs, is an open question.

AI systems. The AI case demonstrated that the framework can analyse AI-mediated harm and benefit across all dimensions. The contested moral-status question remains open: whether AI systems themselves can be bearers of harm on the framework's dimensions (not only producers of harm to others) is a question the framework's method can approach (track observable features, apply precaution) but cannot currently resolve. The connection to the consciousness pa-

per's processing-system account is direct: if the binary conscious/not-conscious collapses into a spectrum of aversive-state capacity and self-modelling depth, the framework's dimensional analysis applies along that spectrum rather than requiring a threshold determination.

Role-based obligations. The Confucian, Hindu, and Islamic traditions all embed ethics in situated roles: parent, ruler, teacher, student, elder. The framework currently treats role as a contextual modifier on the existing dimensions (a parent has different relational obligations than a stranger) rather than as an independent dimension. But the strength and persistence of role-based ethics across independent traditions raises the question of whether situated obligation is tracking something that the current dimensions do not fully capture. If so, the dimensional set may need to be expanded. If not, the framework should explain more precisely how role-dependent variation is generated by the existing dimensions.

What the open questions mean

The open questions stated in this section are not weaknesses to be concealed. They are the natural product of a framework that takes its own diagnostic standard seriously. A framework that claimed to have resolved the minimal basis, the floor-setting problem, the weighting question, the trajectory typology, the distributional complexity, and the scope questions simultaneously would be claiming more than any single document can deliver. The framework's contribution is to identify these questions with precision, to show why they matter, and to indicate where the answers are most likely to come from: empirical investigation for the trajectory typology and dimensional interactions, political and institutional negotiation for the floor-setting and weighting questions, and ongoing philosophical and scientific work for the scope questions. The framework is designed to support all of these lines of investigation. It does not pretend to have completed them.

15. Concluding Observations

This document began with a claim: every decision is a harm profile, not a harm score. The argument that followed has shown that this claim is not a philosophical preference but a structural finding, supported by the independence of the dimensions, the convergence of traditions that detected them independently, the inadequacy of single-axis frameworks when applied to real cases, and the capacity of the multi-dimensional analysis to reveal structure in trajectories that no single-axis framework can register.

The framework that results is not a theory in the traditional sense. It does not propose a single principle from which moral conclusions can be derived. It proposes a set of real variables, a structured method for reasoning about them, a single stated normative presupposition (that structural degradation of entities capable of bearing it is pro tanto bad), and a set of constraints (floors, dominance, temporal analysis) that narrow the space of defensible decisions without collapsing it to a point. It is stronger as a diagnostic for harm than as a description of positive goods, and it acknowledges this honestly rather than concealing it.

The traditions were partial instruments. Each was measuring something real, and each was missing something real. The non-Western traditions that tracked multiple dimensions simultaneously were closer to the full picture than the Western monist traditions that dominated the theoretical conversation. The dimensions the traditions were tracking are more fundamental than the theories built around them, and the tradeoffs between those dimensions are more honest than any single-axis resolution that claims to dissolve them.

The framework connects to the empirical sciences in a way that most ethical frameworks do not. If the dimensions are real and observable, their dynamics over time are empirical phenomena. The nine cases suggest that harm trajectories fall into recognisable types, that escalating multi-dimensional profiles become legible at intermediate stages, and that trajectories are never closed. These are testable claims. The behavioural and systemic sciences already possess the methods for investigating them. The philosophical contribution is the dimensional decomposition, the diagnostic standard, and the normative structure. The empirical investigation of trajectory dynamics is future work, but it is work for which the framework has been designed to provide the foundation.

The framework's dimensions apply to any entity capable of bearing harm on them. The cases demonstrated this for human populations, for ecosystems, for individual sentient animals, and for AI systems whose moral status is uncertain. The method, which tracks observable features and does not require resolution of metaphysical questions before governance can begin, provides a principled approach to moral consideration that extends as far as the phenomena extend, rather than stopping at a boundary drawn by convention or by the limits of a particular tradition's scope.

The plurality of the dimensions is not a weakness to be overcome but a structural feature to be respected. The moral world has more than one axis. The traditions that detected different

axes were not wrong about what they found. They were wrong only when they claimed that their axis was the only one. The most defensible ethics is the one that holds all the axes visible, takes time seriously, extends its analysis to all entities capable of bearing harm on its dimensions, and does not pretend the hard tradeoffs away.

What follows from that is still being worked out. But the direction is clearer than it was at the start.

Working Paper · 2026 · Shaky Foundation Draft 0.2 · MIT Licence