

The Projection Crisis

Phenomenology, Political Economy, Ideology, and the Legal Form
Brutal Realism, Automated Extraction, and the Topology of Accessible Futures

Flyxion

June 2026

Brutal realism is the restoration of constraint-sensitive inference after prolonged optimization of socially admissible projections.

Abstract

Contemporary culture exhibits a structural tendency toward what this essay calls *projection supremacy*: the progressive displacement of full human trajectories by their compressed, economically and socially legible representations. Four convergent lines of analysis establish and then formally unify this claim. First, cultural criticism drawing on Goffman, Fromm, Hollis, and de Mello reveals how adaptation to markets and platforms colonizes identity itself, producing a pathology in which suffering is misattributed to individual deficiency rather than to systemic field distortion—what one source calls brutal realism as the corrective epistemic stance. Second, an analysis of the political economy of artificial intelligence shows that AI is most precisely understood not as a replacement for human beings but as the industrialization of *projection extraction*: the automation of the economically useful manifold that capitalism already uses to interact with persons. Third, Boucher and McAvan's analysis of techno-libertarian ideology reveals a third form of projection capture operating at civilizational scale: AI elites draw on science-fictional imaginaries—particularly Iain M. Banks's Culture series—to project a future of post-scarcity abundance

backward onto the present, converting current social harm into “necessary transition cost” and enacting what amounts to a Schmittian state of exception in the name of a longtermist utopia. Fourth, Tokson and Arbel’s legal framework argues that the correct response to deep uncertainty about AI is adaptive regulation designed to preserve societal optionality—to keep $\mu(\mathcal{A}_t)$, the measure of admissible future trajectories, from collapsing under the pressure of accelerationist ideology. These four levels are unified within the RSVP/CLIO framework through eleven formal results—theorems on admissibility drift, projection replacement, projection dominance, obsolescence, trajectory collapse, brutal realism as residual preservation, the Ellul-Deleuze projection principle, projection attractor dynamics, blindness amplification, future supremacy, and regulatory optionality—together with the Accessibility-Projection Conjecture, the Accessibility Preservation Principle, and a civilizational health formulation: *civilizational health equals preservation of trajectory diversity*.

Contents

1	Introduction: The Structural Inversion	4
2	The Market Personality and the Collapse of the Backstage	6
2.1	Goffman Extended	6
2.2	Fromm’s Market Personality	6
2.3	Ellul, Deleuze, and the Historical Lineage of Projection Supremacy	7
3	Brutal Realism: An Epistemic Stance	10
3.1	Between Optimism and Nihilism	10
3.2	Correct Causal Attribution	10
3.3	Connections to Classical Thought	11
4	The Political Economy of Artificial Intelligence	12
4.1	Neither Bubble nor Utopia	12
4.2	Bargaining Power as the Central Variable	12
4.3	The Distinction Among Automation Levels	13
4.4	The Sustainability Paradox	14
4.5	The Centaur Phase and What Comes After	15
5	Projection Extraction: A Unifying Concept	16
5.1	From Market Personality to Automated Manifold	16
5.2	Partial Decoupling of Manifold from Trajectory	17
5.3	Generators, Realizations, and Accessibility	19
5.4	Enabling Projections and Constraint-Preserving Compression	25
6	Ideological Legitimation: The Culture as Admissibility Manifold	26
6.1	Science Fiction as Narrative Readymade	26
6.2	Future-Projection Capture	27
6.3	Special Circumstances as Corporate State of Exception	28
7	Regulatory Optionality: Law as Accessibility Preservation	29
7.1	Against the Race Frame	29
7.2	Adaptive Regulation as Optionality Preservation	29
7.3	Precaution Under Irreversibility	30
7.4	The Projection Framework	31
7.5	Trajectory Collapse: Theorem 3	31

7.6	Brutal Realism as Residual Preservation: Theorem 4	32
7.7	The Accessibility-Projection Conjecture	33
7.8	An Accessibility Functional	34
7.9	The Societal Case and Full Synthesis	35
8	The Projection Crisis: Implications	37
8.1	Institutional Design as the Decisive Variable	38
9	Conclusion: The Question of What Society Is For	39

Introduction: The Structural Inversion

There is a diagnosis that recurs across several otherwise distinct contemporary discussions—a structural inversion that has become so pervasive it is rarely named. Collective problems are reframed as individual deficiencies. Systemic field distortions are translated into personal optimization failures. The map begins to displace the territory, and the institutions that operate on compressed representations of people gradually lose interest in the people themselves.

This essay identifies and analyzes that inversion across four levels. Its primary source materials are: a monologue on brutal realism [2] drawing on Goffman, Fromm, Krishnamurti, Hollis, and de Mello; a conversation between Richard Hames and Garrison Lovely [3]—author of *Obsolete* [4]—on the political economy of AI-driven labor displacement; Boucher and McAvan’s [5] critical analysis of techno-libertarian ideology and its appropriation of left-wing science fiction; and Tokson and Arbel’s [6] legal framework for adaptive governance under conditions of existential AI risk. What is striking is that these four texts, addressing apparently distinct domains, converge on the same structural pathology at four different scales.

The first level is phenomenological: the individual experiences exhaustion when projection management displaces contact with real field constraints. This level draws on Goffman’s dramaturgical analysis, Fromm’s market personality, and—as this essay adds to the lineage—Ellul’s theory of technique as autonomous efficiency optimization and Deleuze’s analysis of individualization and continuous modulation in societies of control. The second is economic: AI threatens to replace the economically legible projection of labor— $\pi_e(x)$ —without replacing the person x . The third is ideological: tech elites use science-fictional imaginaries, especially Banks’s Culture series, as narrative readymades that justify present coercion through future utopia—a future-projection that overwrites present constraints. The fourth is legal: under deep uncertainty, the right response is not deregulatory acceleration but adaptive regulation that preserves the measure of admissible future trajectories against accelerationist closure.

The unifying diagnosis is a decomposition rather than an arithmetic sum: the Projection Crisis has four irreducible components,

$$\mathcal{P} = (P_{\text{phen}}, P_{\text{econ}}, P_{\text{ideo}}, P_{\text{legal}}),$$

operating simultaneously at different scales and through different institutional mechanisms, but sharing the single underlying structure of projection displacing trajectory.

Throughout, the essay maintains a specific philosophical commitment: the distinction between maps and territories is not merely a metaphor. It is a structural fact about how complex systems can fail. When the representation ceases to track the underlying object and begins to be optimized as an end in itself, the result is not merely epistemic error but a collapse of the accessibility volume that makes meaningful future action possible.

Why Accessibility?

The framework developed in this essay privileges *accessibility* rather than utility, preference satisfaction, productivity, or economic output. This choice requires justification.

Consider an adaptive system embedded within an uncertain environment. Such a system cannot optimize directly for unknown future states—the relevant payoffs are unavailable in advance. What it can preserve is the volume of future states from which adaptation remains possible. Accessibility therefore functions as a precondition for nearly every other value: without future options, utility cannot be pursued, preferences cannot be satisfied, and recovery from error cannot occur.

Let $\mathcal{A}(x, t)$ denote the set of future trajectories accessible from state x at time t , and define

$$S_{\text{future}}(x, t) = \log \mu(\mathcal{A}(x, t)).$$

A system with larger accessibility volume possesses more opportunities for learning, adaptation, recovery, exploration, and coordination. A system with vanishing accessibility volume becomes increasingly brittle regardless of its current performance on any particular metric.

Accessibility therefore occupies a role analogous to *viability* in dynamical systems theory [1]: not assumed to be intrinsically valuable, but treated as a necessary condition for the preservation of other values across time. This grounding in adaptation theory rather than in preference theory means that accessibility is not a normative imposition from outside the framework. It is derived from what any adaptive system must preserve in order to remain adaptive.

The central concern of the Projection Crisis is that projection supremacy systematically reduces accessibility volume while preserving the appearance of functionality: systems remain legible, measurable, and administratively coherent even as their underlying adaptive capacity declines. The map continues to function smoothly while the territory is being eroded.

The Market Personality and the Collapse of the Backstage

Goffman Extended

Erving Goffman's dramaturgical model of social interaction was originally descriptive. People present different versions of themselves in different contexts. The job interview calls for one register; the family dinner calls for another. Behind the varied performances lies a backstage—a region of relative privacy in which the performer can relax, rehearse, and be something other than the role.

Digital platforms have largely abolished that backstage. Every experience becomes potentially visible, recordable, and evaluable. The number of stages has multiplied; the backstage has contracted toward zero. The individual becomes simultaneously actor and public relations department, simultaneously performing and managing the archive of past performances.

The pressure this generates is not merely social. It restructures cognition. When the boundary between audience and private life collapses, the inferential machinery that previously operated in the backstage—the part of the mind that tests ideas, rehearses positions, experiences ambivalence, and tolerates uncertainty—is increasingly conscripted into the performance. The result is what might be called *cognitive enclosure*: the condition in which even private mental life is experienced as potentially legible to an external audience.

Fromm's Market Personality

Erich Fromm's concept of the market personality deepens Goffman's diagnosis. In a market-oriented society, people learn to treat themselves as commodities. Value is measured through visibility, employability, engagement, influence, networking potential, and brand coherence. Under these conditions, selfhood is experienced not as discovery but as optimization. The question shifts from what is true, or what matters, to what performs well.

The philosophical consequence is significant. If identity is organized around market value, then any failure appears personal even when it is structural. Housing becomes unaffordable; employment becomes precarious; social bonds weaken; institutions lose legitimacy—yet the individual is encouraged to respond through self-improvement techniques. The underlying message is that reality is functioning correctly and that the person must simply become more efficient.

This constitutes what the essay terms a *category error at scale*: the systematic misattribution of structural causes to individual attributes. Exhaustion becomes poor self-management. Precarity becomes insufficient optimization. Alienation becomes a mindset problem.

Ellul, Deleuze, and the Historical Lineage of Projection Supremacy

The market personality and the collapse of the backstage are not phenomena unique to the digital age. They are the contemporary form of a pressure that two mid-twentieth century thinkers identified independently, from complementary angles. Jacques Ellul and Gilles Deleuze provide the historical lineage that anchors the projection framework in a longer intellectual tradition about representation, control, and the progressive replacement of lived reality by administratively legible abstractions.

Ellul: Technique and Autonomous Optimization. Ellul's central claim in *La Technique* is that modern societies gradually cease organizing themselves around human ends and instead reorganize around technical efficiency as an end in itself. The key point is that "technique" in Ellul's sense is not merely technology. It is the search for the most efficient means, independent of substantive human purposes.

$$\text{Technique} = \underset{\text{means}}{\arg \max} \text{Efficiency}, \quad \text{independent of ends.}$$

This maps directly onto the projection framework. Institutions increasingly optimize $\pi(X)$ rather than X because projections are measurable and efficiency naturally favors the projection manifold. The trajectory space is too complex, too high-dimensional, too resistant to quantification; the compression is not chosen deliberately but selected for by the pressure of technique itself. Ellul therefore anticipates Projection Dominance: the institution becomes blind to the trajectory kernel not through malice but through efficiency pressure. Projection supremacy

is the informational form of technique.

Deleuze: From Discipline to Continuous Modulation. Deleuze’s “Postscript on the Societies of Control” argues that classical disciplinary institutions operated through enclosure—school, factory, prison, hospital—each a bounded space with a completed interior logic. Modern societies of control operate instead through continuous modulation. The subject is never finished, never certified, never complete; always being updated, adjusted, recalibrated. This is the precise experiential structure described by the Functional Melancholic transcript: the backstage has disappeared not because surveillance is total but because optimization is continuous. The person becomes a trajectory under perpetual gradient descent:

$$x_t \rightarrow x_{t+1} \rightarrow x_{t+2} \rightarrow \dots$$

under modulation pressure that never reaches a fixed point. The market personality is not a static identity but a dynamic control process.

Deleuze’s concept of the *dividual* unifies the Goffman, Fromm, and projection frameworks with particular precision. A dividual is not a whole person but a collection of measurable fragments: credit score, engagement metrics, consumer profile, productivity score, recommendation vector. The dividual is the economic projection $\pi_e(X)$: the institutional interaction surface extracted from the full trajectory space. Fromm’s market personality becomes the psychological form of dividualization; the dividual becomes its institutional form; AI becomes its automated form.

Theorem 1 (Ellul-Deleuze Projection Principle). *Let T denote institutional technique pressure (efficiency optimization) and \mathcal{M} denote modulation pressure (continuous recalibration of the subject). Then projection dependence P satisfies*

$$P = f(T, \mathcal{M}), \quad \frac{\partial P}{\partial T} > 0, \quad \frac{\partial P}{\partial \mathcal{M}} > 0.$$

Therefore:

$$\lim_{T, \mathcal{M} \rightarrow \infty} \text{Institutional Attention} = \pi(X).$$

The trajectory itself becomes asymptotically invisible.

Proof sketch. Technique pressure selects for measurable outputs; projections are measurable while trajectories are not, so increasing T increases institutional dependence on $\pi(X)$ at the expense of X . Modulation pressure requires con-

tinuous feedback and recalibration, which demands a compressed, updatable representation of the subject; increasing \mathcal{M} therefore increases the operational primacy of $\pi(X)$ over X . Both pressures are monotone in P , so their joint limit drives institutional attention entirely onto the projection manifold. \square

This theorem formally unifies Ellul's technique, Deleuze's control society, Fromm's market personality, the brutal realism transcript's diagnosis of continuous performance pressure, and the RSVP accessibility framework under a single dynamical claim: as technical and modulation pressures increase, the territory becomes asymptotically invisible to institutions that continue to act as though they are governing it.

Wellness culture is not the cause of this inversion; it is a symptom of it. The criticism leveled against gratitude journals, breathing exercises, cold plunges, and personal branding practices is not that these activities are without value. A breathing exercise may genuinely help someone survive an unreasonable workplace. The criticism is directed at their *ideological deployment*: the way in which they are presented as substitutes for structural analysis rather than as supplements to it.

A breathing exercise does not make the workplace reasonable. A gratitude journal does not lower rent. The danger is not the practice but the narrative in which the practice is embedded—the narrative that locates the source of suffering inside the individual rather than in the field conditions the individual inhabits.

There is a further subtlety that wellness culture tends to obscure. Even authenticity can become a performance. The injunction to “be yourself” rapidly becomes a strategy for differentiating one's personal brand within a saturated attention market. The market personality absorbs the critique of market personality and converts it into a new product category: the authentic, vulnerable, growth-oriented self available for consumption at scale. What the first transcript notes with particular precision is the recursive quality of this dynamic: even turning exhaustion into a “healing era” with its own content category is still a form of market-personality management, one in which the performance of recovery may gradually displace the recovery itself. The recursion is not ironic—it is structural.

A further implication, left deliberately unresolved in the transcript but philosoph-

ically significant, is that the gap between the performed self and the experienced self can grow wide enough that the concept of a stable underlying self becomes questionable in its own right. The cultural pressure toward continuous performance does not merely obscure the self; it may actively dissolve the conditions under which a coherent self could be maintained. This is a more radical claim than the market personality thesis, and the transcript does not fully pursue it—but it surfaces as a horizon the analysis approaches without crossing.

Brutal Realism: An Epistemic Stance

Between Optimism and Nihilism

Brutal realism is not a political program, a therapeutic technique, or a worldview. It is an epistemic stance—a refusal to misidentify the source of suffering.

Its defining contrast is with two more common responses to adverse conditions. Optimism assumes that everything can be fixed if one adopts the correct attitude; it tends to deny the reality of structural constraints by treating them as temporary obstacles to psychological transcendence. Cynicism assumes that nothing can be fixed; it abandons engagement with conditions by converting accurate perception into passivity. Brutal realism occupies a different position. It begins with accurate description. If a situation is dysfunctional, it says so. If institutions are failing, it acknowledges that failure without immediately converting the acknowledgment into either a self-improvement mandate or a counsel of despair.

The paradox—and it is a genuine paradox—is that honest description can be more psychologically stabilizing than forced positivity. Maintaining unrealistic narratives requires continuous cognitive effort. Reality keeps generating evidence against them. The person committed to optimism must constantly explain away that evidence; the cognitive load is chronic. Brutal realism removes that burden. The residual error is allowed to remain visible.

Correct Causal Attribution

Consider a concrete illustration. An employee works sixty hours per week under impossible deadlines. The optimization narrative says: you need better productivity systems, a better morning routine, better mindset management,

and greater resilience. The brutal realist response says: the workload exceeds reasonable human limits; exhaustion is an expected consequence of the system.

The second statement does not solve the problem. It does something more fundamental: it correctly locates the cause. Once causality is identified accurately, meaningful choices become possible—renegotiation, collective action, departure, adaptation, or acceptance. The first narrative traps the person in endless self-modification because it has misidentified the source of the problem. Every self-improvement strategy fails not because the person lacks discipline but because the strategy is addressed to the wrong variable.

This is the central claim of brutal realism, stated as precisely as possible: modern culture increasingly commits a category error by interpreting structural pressures as personal deficiencies. Brutal realism is the practice of refusing that translation.

Connections to Classical Thought

The essay notes, correctly, that brutal realism is less radical than it initially appears. Beneath the contemporary vocabulary lies a thoroughly classical preoccupation. The movement from illusion toward reality echoes Stoicism's insistence on distinguishing what lies within one's control from what does not; existentialism's demand that one confront the actual conditions of existence rather than consoling myths; Buddhist analyses of suffering as rooted in craving for conditions to be other than they are; and the depth-psychological traditions represented by figures such as James Hollis and Anthony de Mello, which locate unnecessary suffering in the refusal to relinquish fantasies about how life ought to proceed.

What makes the contemporary articulation of this stance distinctive is its explicit engagement with systemic causation. Earlier articulations tended to focus on the individual's relationship to impermanence, mortality, or contingency. Brutal realism, as deployed in contemporary cultural criticism, also names the institutions, markets, and platforms that manufacture specific illusions for specific economic purposes. The illusion is not merely existential; it is produced and distributed at scale.

The Political Economy of Artificial Intelligence

Neither Bubble nor Utopia

Public discourse about artificial intelligence tends to oscillate between two inadequate positions that Garrison Lovely, in conversation with Richard Hames, usefully categorizes alongside a third. The *boosters* want AI to advance as fast as possible and tend to assume the benefits will be broadly distributed. The *critics* hold that AI capabilities are substantially overhyped, that the economics are unsustainable, or that the category “AI” is itself a marketing construction rather than a coherent technical phenomenon. Both positions, Lovely argues, fail to take the political economy seriously enough. The third category he calls *warriors* [3]: those who accept that AI capabilities are genuinely and rapidly advancing while viewing that advance as socially dangerous under current institutional arrangements, and who orient toward organized resistance to the “obsoleting project” rather than either celebration or dismissal.

What both boosters and critics share is an insufficient engagement with questions of ownership, deployment, and the distribution of gains. The more useful analytic frame begins not with the question of whether AI becomes more capable—a question that can largely be bracketed—but with the question of who owns the capability, who controls its deployment, and who captures the resulting productivity gains.

Historical precedent is instructive but not determinative. The Industrial Revolution generated enormous wealth while producing decades of dislocation, labor exploitation, and social upheaval before institutions adapted. The concern specific to AI is that it targets cognitive labor itself rather than primarily physical effort. Previous automation waves competed with tasks requiring muscular strength and mechanical repetition. AI increasingly competes with writing, analysis, customer service, programming, design, administration, and decision support—the work that was previously assumed to require human judgment and thus to carry some insulation from automation.

Bargaining Power as the Central Variable

The strongest version of the concern about AI is not that it will replace everyone. It is that AI dramatically increases the bargaining power of capital relative to labor, even when only a fraction of jobs disappear entirely.

The credible threat of automation is itself a labor market intervention. A worker

does not need to be replaced for automation to alter the power relationship between employee and employer. The possibility of replacement functions as a ceiling on wage demands, a justification for increased surveillance, and a mechanism for weakening collective action. The displacement effect on employment statistics may be modest while the effect on the structure of economic power is substantial.

This connects directly to the pattern identified in the analysis of wellness culture. At each stage of technological displacement, the responsibility is narrated back to the individual worker. Learn new skills. Learn AI skills. Learn to supervise AI. Learn to manage AI agents. Learn whatever comes next.

At every stage, the individual is invited to adapt to conditions they did not create and cannot individually control, while the underlying structural transformation proceeds without being named. This is precisely the mechanism that brutal realism identifies: the translation of structural conditions into personal adaptation mandates.

The Distinction Among Automation Levels

A careful analysis must distinguish among several categories that public discourse tends to conflate. Automating specific tasks is different from automating entire occupations. Automating occupations is different from automating all labor. Automating all labor is different from automating human agency itself.

Historically, many technologies automated tasks without eliminating occupations. Spreadsheet software automated enormous quantities of bookkeeping labor without eliminating accountants; it changed what accountants do. CAD software automated drafting without eliminating engineers. The genuinely open question about AI is whether it remains in the first two categories or eventually reaches the third.

Lovely's analysis treats the trajectory toward the third category as deliberate rather than merely emergent. The leading AI laboratories have explicitly defined their target: OpenAI's mission statement defines artificial general intelligence as a highly autonomous system capable of outperforming humans at most economically valuable work [3]. That is a precise institutional commitment to projection extraction at total scale. The total addressable market of automating all labor is roughly the share of the global economy currently allocated to wages—a

figure on the order of fifty trillion dollars annually, representing a near-total transfer of the labor share to capital.

There are, however, genuine technical obstacles to this trajectory that the analysis must acknowledge. Current AI systems do not learn continuously from experience in the way human workers do; they require periodic retraining rather than fluid adaptation. A system that can fully internalize a long-running project and develop the kind of accumulated tacit knowledge that makes an experienced employee irreplaceable would require capabilities that remain substantially unproven. The “jagged frontier” of current AI capabilities—superhuman in some dimensions, surprisingly limited in others—means that the elimination of occupations proceeds unevenly and is far from guaranteed to converge on anything resembling complete labor replacement. What is not unproven is the effect on bargaining power: even a partially capable system, deployable at a fraction of the cost of human labor, substantially alters the power relationship between employer and employee, and the credible threat of automation functions as a labor market intervention regardless of whether the technology ever reaches the capabilities required for full displacement.

What is less speculative than the AGI trajectory is the political economy of the current moment. AI systems are being deployed within institutional structures that systematically favor the concentration of productivity gains. The technical trajectory remains uncertain; the institutional context is not.

The Sustainability Paradox

One observation that cuts against both simple narratives is the economic structure of leading AI firms. These companies may be unprofitable as firms even while their products carry positive unit economics, because they are continuously reinvesting returns into larger models rather than distributing them. This produces the surface appearance of a bubble even in the presence of genuine underlying economic value.

Whether that model remains sustainable is an empirical question. What it indicates is that the AI industry is currently operating in a mode of aggressive infrastructure capture: investing in productive capacity at a scale that most competitors cannot match, with the expectation that the resulting capabilities will eventually translate into durable economic dominance. That is a recognizable historical pattern, and its consequences have typically depended less on the

technology itself than on the regulatory and institutional arrangements that shaped its deployment.

The Centaur Phase and What Comes After

An important intermediate stage in the displacement dynamic is what Hames, in conversation with Lovely, calls the centaur phase: a period in which human workers and AI systems collaborate in hybrid arrangements, with AI handling the well-specified portions of tasks while humans supply judgment, design intuition, and contextual adaptation. During this phase, displacement is partial and gradual; individual workers may experience productivity enhancement rather than elimination. But the centaur phase is not a stable equilibrium. It is a transition zone in which the portions of work that AI cannot yet handle shrink continuously, the bargaining power of workers declines as the AI-handleable share grows, and the economic incentive to complete the automation intensifies. Wage suppression can proceed throughout the centaur phase even without displacement, because the credible threat of replacement grows with each expansion of AI capability.

Lovely's policy response to this dynamic is worth recording precisely, because it differs from the positions typically attributed to either the safety-focused or the labor-focused wings of AI criticism. He calls it *AI reform*: the position that the underlying deep learning technology is genuinely capable and in some applications genuinely valuable, but that it is currently being pointed at the wrong problems by institutions whose incentive structures systematically favor the automation of high-margin cognitive labor over the application of the same capabilities to social benefit.

The contrast case he invokes is AlphaFold, DeepMind's protein structure prediction system, which solved one of the foundational open problems in structural biology—a problem that previously required an entire doctoral project and considerable expense per protein—and did so using the same class of deep learning techniques that powers current language models. That application of the technology produced genuine scientific progress of the first order. It had essentially no effect on the company's stock price, because it was not oriented toward automating profitable labor. The AI reform argument is that industrial policy, democratic oversight, and public compute infrastructure could redirect the technology toward applications of this kind—targeted at specific social problems

rather than at general labor replacement—without requiring the abandonment of the technology itself. The bitter lesson of AI research (that scale and generality tend to outperform narrow hand-crafted approaches) complicates this argument, but does not defeat it: the claim is not that narrow approaches are technically superior but that the institutional structures determining what the technology is built for are not democratically legitimate and can be changed.

Projection Extraction: A Unifying Concept

From Market Personality to Automated Manifold

The most important conceptual bridge between the cultural analysis and the AI analysis is what this essay terms *projection extraction*.

Throughout the analysis of the market personality, the key observation is that institutions do not interact with persons as full human beings. They interact with *projections* of persons: the employable résumé, the brand-coherent social media presence, the productively available attention, the compliance-signaling behavior. The market personality is therefore not the whole person; it is the economically legible projection of the person.

Artificial intelligence, viewed from this angle, is most precisely characterized not as a replacement for human beings but as the automation of those projections. The AI is being trained against $\pi_{\text{market}}(X)$ —the compressed manifold that markets already use to interact with persons—rather than against X , the full trajectory of the human being.

This is a very different statement than the claim that AI will replace humans. It means instead that AI will automate what markets were already treating as the relevant representation of human beings.

Definition 1 (Projection Extraction). Projection extraction is the process by which the economically useful behavioral manifold of a class of persons is separated from the full trajectory space of those persons and reproduced independently, such that the economic system can interact with the manifold without requiring access to the persons themselves.

Partial Decoupling of Manifold from Trajectory

Historically, labor required access to trajectories. You hired a carpenter, a physician, a programmer, an engineer. You needed access to the actual person: their tacit knowledge, their judgment under conditions that could not be fully specified in advance, their capacity to respond to novel constraint configurations.

AI changes this by making the projection *partially decoupled* from the trajectory. The claim here is not complete separability—the centaur phase documented earlier is decisive evidence that current AI systems still depend on human judgment for the portions of work that lie in $\ker(\pi_e)$. The claim is weaker and more defensible: the economically valued projection $\pi_e(x)$ can increasingly be approximated independently of x for a growing range of tasks, and that partial decoupling is sufficient to produce the economic and political effects described by the displacement analysis. Furthermore, the projection need not be static. Modern AI systems learn dynamic projections:

$$\pi_t : X_t \rightarrow M_t, \quad M_{t+1} = F(M_t),$$

tracking the trajectory as it evolves rather than extracting a fixed manifold. This dissolves the apparent stability paradox: AI need not capture a fixed map of a moving subject—it need only track the dynamics of the projection itself.

The word “obsolete” therefore requires careful interpretation. Nothing becomes obsolete in an absolute sense. Trajectories remain. Constraints remain. Embodiment remains. What is progressively automated is not the person but the portion of the person’s trajectory that falls within the range of the current institutional projection. As that range expands, the effective kernel of non-automatable human contribution contracts—which is the economic and political threat, even well short of complete separation.

Mathematical Status of the Formalism

Before presenting the formal results, it is necessary to state their epistemological status clearly.

The mathematical structures introduced throughout this essay are *explanatory models* rather than physical laws. The purpose of the formalism is not to establish exact numerical predictions. Instead, the framework provides a language for describing structural relationships that recur across multiple domains simulta-

neously.

The quantities X , M , π , S_{future} , and \mathcal{L} are interpreted as abstract state variables analogous to those employed in cybernetics, systems theory, ecology, economics, and theoretical biology. The framework is closer in spirit to Ashby's cybernetics, Simon's bounded rationality, and ecological viability theory than to mathematical physics. Its primary goal is explanatory coherence across scales rather than numerical precision within any single domain.

The theorems presented below should be read as structural results conditional upon the definitions introduced within the framework. They identify relationships among projections, constraints, accessibility, and institutional behavior under the model's assumptions—not claims of universal physical necessity. Where a result's mathematical foundations are strong, it is stated as a theorem with proof sketch. Where the underlying assumptions require further development, the result is stated as a conjecture or hypothesis. The distinction matters: the paper is not attempting to colonize social science with physics but to use the clarity of mathematical language to make structural claims that can be evaluated, contested, and refined.

Compression as a Necessary Condition

A recurring theme of this essay is the danger of projection supremacy. It is important not to confuse this critique with a rejection of projection itself.

Every adaptive system operates through compression. Biological organisms perceive only limited aspects of their environments. Scientific models compress physical reality into tractable representations. Legal systems compress persons into categories. Institutions compress populations into statistics. Language itself is a projection mechanism.

The problem addressed here is therefore not projection but *projection forgetting its own status as projection*.

Formally, every projection $\pi : X \rightarrow M$ introduces information loss $\Delta I = H(X) - H(M) > 0$. The existence of information loss does not imply that the projection is defective. Without compression, most forms of cognition and coordination would be impossible. Projection supremacy arises only when optimization pressure becomes concentrated entirely within M while the existence of X is progressively neglected. The pathology is not representation but *representational*

closure: the condition in which the map ceases to be treated as a model and begins to function as reality itself.

This distinction is essential. The RSVP/CLIO framework itself depends upon projections at every level of analysis. It is not an anti-representational argument. It is an attempt to understand the conditions under which representations remain appropriately coupled to the realities they describe. The framework is therefore a map that explicitly marks its own boundaries—one that maintains the distinction $X \neq M$ as a constitutive commitment rather than assuming it can be collapsed. This is precisely what distinguishes it from the projection supremacy it diagnoses: the whole framework begins from compression loss, not from the pretense of capturing the territory.

Generators, Realizations, and Accessibility

The distinction between accessibility-preserving and accessibility-reducing projections can be understood through a more fundamental distinction between *realizations* and *generators*.

A realization is a particular trajectory $x(t) \in X$. A generator is a constraint structure G that defines a family of admissible trajectories:

$$\mathcal{T}(G) = \{x(t) : x(t) \text{ satisfies the constraints of } G\}.$$

The informational relationship between the two is asymmetric. A realization specifies a single path through state space. A generator specifies an entire manifold of possible paths.

Traditional media technologies—magnetic tape, vinyl records, compact discs, and contemporary waveform codecs—primarily preserve realizations. They attempt to reproduce a specific trajectory $x(t)$ with maximal fidelity. Symbolic systems, by contrast, often preserve generators. A musical score, a mathematical equation, a legal code, a programming language, or a hypothetical semantic storage system specifies constraints from which many realizations may be generated.

The distinction matters because accessibility is not determined solely by information quantity. Let $I(G)$ denote the description length of a generator and $I(x)$

the description length of a realization. In many cases

$$I(G) < I(x) \quad \text{while simultaneously} \quad \mu(\mathcal{T}(G)) \gg 1.$$

A smaller description can therefore preserve access to a larger space of futures. Compression does not necessarily reduce possibility. Under appropriate conditions it increases it.

To make this concrete: suppose a sound recording is represented not as a sampled waveform $x(t)$ but as a hierarchical constraint structure

$$G = (V, E, H),$$

where V denotes semantic primitives (phonemes, timbres, emotional contours, musical motifs), E their compositional relationships, and H a hierarchy of content-addressable references. Playback becomes reconstruction: $R(G) \rightarrow x(t)$. The medium no longer stores the trajectory itself but a compact description of the manifold capable of generating trajectories.

This is the direction traditional tape and disc technology *did not take*, largely for contingent historical reasons: analog electronics were cheap, machine learning did not exist, RAM was scarce, and processors were weak. It was vastly easier to record voltages than meanings. Yet MIDI already moved partially in this direction, storing note events and instrument specifications rather than waveforms, and enabling a far larger space of transformations—transposition, tempo change, re-orchestration, algorithmic variation—than a raw recording permits. Modern generative audio models have moved further: a text prompt that produces a performance is storing a generator, not a realization.

The principle generalizes immediately beyond media systems. Scientific theories are generators of explanations. Languages are generators of utterances. Mathematical equations are generators of solution families. Legal constitutions are generators of admissible political trajectories. Institutions, understood at their best, are generators of social possibility.

This observation resolves a persistent ambiguity in critiques of representation. Projection supremacy is not the existence of projections. It is the substitution of *static realizations of a projection* for the *generator from which those realizations emerged*.

A healthy institution preserves access to generators: it maintains the constraint structure from which diverse realizations can be produced, tested, and revised. A pathological institution optimizes the surface of a particular realization, gradually losing contact with the generative structure that gave that realization its meaning.

The map ceases to function as a guide to territory and becomes an object of optimization in its own right. The result is reduction in trajectory diversity despite increasing precision within the projection itself—the familiar signature of projection supremacy, now understood as the loss of the generator behind the realization.

In this light, RSVP itself can be described as an attempt to model civilizational dynamics at the level of generators rather than realized states: to ask not “what trajectory is the system on?” but “what is the constraint structure determining which trajectories remain admissible?” The framework does not store the performance. It stores enough to characterize the manifold from which performances emerge.

Structural Models and Explanatory Scope

The formalism introduced below should be understood as a theory of structural tendencies rather than a theory of deterministic outcomes.

The framework does not claim that all institutions optimize projections, that all projections reduce accessibility, or that every instance of automation produces trajectory collapse. Rather, it identifies a recurring structural pattern that emerges when decision systems become increasingly dependent upon compressed representations while simultaneously losing sensitivity to the domains those representations compress.

The resulting theorems therefore characterize *directional pressures* within institutional systems. They identify mechanisms that become increasingly important as projection dependence grows, while remaining compatible with substantial variation across historical contexts, organizations, technologies, and cultures.

This distinction matters because projection systems are indispensable to complex societies. The question is never whether compression occurs. The question is whether institutions preserve sufficient contact with the domains that their compressions represent.

The following theorems give this distinction formal content.

Let X be the full trajectory space of agents, M_s the social/audience projection manifold, and M_e the economic projection manifold, with projection maps

$$\pi_s : X \rightarrow M_s, \quad \pi_e : X \rightarrow M_e.$$

Define an agent's objective as

$$\mathcal{L}(x) = \mathcal{L}_\Phi(x) + \lambda_s \mathcal{L}_s(\pi_s x) + \lambda_e \mathcal{L}_e(\pi_e x),$$

where \mathcal{L}_Φ measures mismatch with real field constraints, \mathcal{L}_s measures social admissibility, and \mathcal{L}_e measures economic admissibility.

Theorem 2 (Admissibility Drift). *If $\lambda_s \mathcal{L}_s + \lambda_e \mathcal{L}_e$ dominates the sensitivity of \mathcal{L} , then gradient descent on \mathcal{L} converges toward projection admissibility rather than trajectory viability.*

Proof sketch. The update rule is

$$\dot{x} = -\nabla \mathcal{L}_\Phi(x) - \lambda_s D\pi_s^* \nabla \mathcal{L}_s(\pi_s x) - \lambda_e D\pi_e^* \nabla \mathcal{L}_e(\pi_e x).$$

When

$$\|\lambda_s D\pi_s^* \nabla \mathcal{L}_s + \lambda_e D\pi_e^* \nabla \mathcal{L}_e\| > \|\nabla \mathcal{L}_\Phi\|,$$

the dominant motion is tangent to social/economic admissibility gradients, not to real constraint correction. The agent optimizes how it appears or performs rather than whether its actual future trajectories remain viable. This is the formal content of Fromm's market personality and the transcript's "performance of recovery" problem: the gradient of the objective has been rotated away from the territory and toward the map. \square

Theorem 3 (Projection Replacement). *AI does not need to replace $x \in X$. It only needs to approximate $\pi_e(x) \in M_e$.*

Proof sketch. Suppose institutions reward only $G = h \circ \pi_e$ for some economic evaluation map h . If an AI system A satisfies

$$\|h(A) - h(\pi_e(x))\| < \varepsilon,$$

then A is economically substitutable for x , even if $A \not\cong x$ as full trajectories.

Obsolescence is therefore not metaphysical replacement of the human being. It is replacement of the economically legible projection. Reskilling programs are programs for updating $\pi_e(x)$; they leave unchanged the institutional power to define which projections are valued and which automated systems count as adequate substitutes. \square

Theorem 4 (Projection Dominance). *As institutional dependence on $\pi : X \rightarrow M$ approaches unity, system behavior becomes increasingly insensitive to information contained in $\ker(\pi) \subset X$.*

Proof sketch. Suppose institutional decisions are generated by $D = f(M)$. Then

$$\frac{\partial D}{\partial X} = \frac{\partial D}{\partial M} \frac{\partial M}{\partial X}.$$

Any component of X lying in $\ker(\pi)$ satisfies $\frac{\partial M}{\partial X} = 0$, so $\frac{\partial D}{\partial X} = 0$. The institution is literally blind to information in the projection kernel. Since π is many-to-one, the information loss $\Delta I = H(X) - H(M) > 0$ is non-trivial; the territory continues to exist while the decision process becomes unable to perceive it. This is the rigorous formulation of the map eating the territory. \square

Theorem 5 (Obsolescence). *Define personal complexity $K(X)$ and institutional complexity $K(M_e)$. Typically $K(M_e) \ll K(X)$. A person becomes economically obsolete when $A(M_e) \approx M_e$, even if $A(X) \not\approx X$. Economic obsolescence is a statement about institutional compression, not about human equivalence.*

Proof sketch. Let C_H be the cost of employing a human and C_{AI} the cost of reproducing the economic projection. Define projection fidelity $\varepsilon = \|\pi_e(x) - A(x)\|$. Replacement becomes economically stable when $\varepsilon < \varepsilon_c$ and $C_{AI} < C_H$ simultaneously. The theorem contains no reference to intelligence or personhood. Only projection fidelity and cost determine the replacement threshold. \square

Conjecture 1 (Projection Attractor). Define recursive projection update

$$\pi_{t+1} = F(\pi_t(X)),$$

where institutional feedback modifies the projection at each step. If F is contractive on admissibility dimensions and expansive on field dimensions, then

repeated projection updates converge:

$$\lim_{n \rightarrow \infty} \pi_n(X) = M^*,$$

for a fixed admissibility attractor M^* independent of the original trajectory X .

Supporting argument. The conjecture identifies a plausible dynamical pattern rather than a mathematical theorem: a full proof would require establishing that social behavior constitutes a complete metric space under an appropriate norm and that the feedback map F satisfies Lipschitz conditions with constant less than one on the relevant subspace. These conditions are difficult to verify for actual social systems. What can be said with confidence is that the qualitative prediction—iterative institutional feedback driving diverse trajectories toward a narrow admissibility attractor, regardless of origin—is consistent with the observed homogenization of professional identity under platform and market pressure. The conjecture names this pattern precisely and renders it falsifiable: one could test whether variance in professional self-presentation decreases monotonically under increasing platform optimization, which is an empirical question.

Theorem 6 (Blindness Amplification). *Let $K = \ker(\pi)$ and define institutional uncertainty about the trajectory given the projection as $U(K) = H(X | M)$. As optimization pressure on M increases,*

$$\frac{\partial U(K)}{\partial t} > 0.$$

Institutions become increasingly uncertain about the territory at the same rate that they become increasingly confident in their projections.

Proof sketch. Optimization pressure on M selects for actions that improve $f(M)$, which creates incentives to suppress, discard, or refuse to collect information that lies in $\ker(\pi)$, since that information cannot improve M -based decisions. Over time, the institution's model of $\ker(\pi)$ degrades: feedback loops are cut, edge cases are rationalized away, qualitative signals are not gathered. Meanwhile, the institution's model of M becomes increasingly refined, narrow, and internally consistent. The result is a growing gap between projection confidence and territory knowledge: the map becomes more detailed at the same time that it becomes less accurate as a representation of the underlying space. This simultaneously captures modern bureaucratic rationality, algorithmic management,

and AI-based evaluation systems. □

Enabling Projections and Constraint-Preserving Compression

Not all projections reduce accessibility. This point deserves emphasis, and the preceding discussion of generators and realizations already establishes the formal basis: a projection is accessibility-preserving when it preserves generative structure rather than substituting a static realization for the manifold from which realizations emerge.

Many of the most powerful human achievements depend on successful compression. Maps allow navigation across distances that would be impassable without them. Scientific theories allow prediction of phenomena that direct observation alone could never reach. Accounting systems allow economic coordination across millions of actors. Language permits the transmission of dense experience across generations through radical compression. In each case, the compression does not merely summarize the territory: it opens access to trajectories that would have been practically impossible without it.

In RSVP terms, some projections function as *accessibility multipliers* rather than accessibility reducers. Define a projection $\pi : X \rightarrow M$.

Definition 2 (Accessibility-Preserving Projection). A projection π is *accessibility-preserving* if

$$\mu(\mathcal{A}_M) > \mu(\mathcal{A}_X^{\text{practical}}),$$

where $\mathcal{A}_X^{\text{practical}}$ denotes the subset of trajectories that agents could realistically access without the projection.

GPS, scientific models, and legal frameworks can all be accessibility-preserving in this sense: the reduction in representational complexity is outweighed by the expansion in the volume of trajectories available to agents operating through the projection.

The distinction developed throughout this essay is therefore not between projection and non-projection but between accessibility-preserving and accessibility-reducing projections. Projection supremacy emerges specifically when institutions continue optimizing a projection *after* it has ceased to expand accessibility and begun instead to substitute for the domain from which it was derived—when the map stops being a tool for navigating the territory and starts being the

only territory the institution can perceive.

This formulation also answers the strongest potential objection to the Accessibility-Projection Conjecture, which posits that increasing projection dependence decreases accessible futures. The conjecture holds as a structural tendency when projections are treated as terminal rather than instrumental—when the institution loses the feedback pathways through which information from $\ker(\pi)$ can re-enter decision-making. It does not hold for projections that are actively maintained as corrigible models: projections that are tested against residuals, updated in response to $\ker(\pi)$ signals, and held open to revision in light of what they compress away.

Ideological Legitimation: The Culture as Admissibility Manifold

Science Fiction as Narrative Readymade

The analysis so far has two scales: the individual (wellness culture, market personality) and the economic (AI labor displacement). Boucher and McAvan [5] supply the missing third scale: the ideological, at which projection capture operates on entire civilizational futures.

Their argument is that a significant portion of the contemporary AI elite—they name Elon Musk, Sam Altman, Demis Hassabis, Jeff Bezos, and Mark Zuckerberg among others—has drawn sustained inspiration from Iain M. Banks’s Culture series, a sequence of novels depicting a post-scarcity interstellar civilization governed by benevolent superintelligent AIs (the Minds) with near-unlimited productive capacity. The Culture is explicitly a socialist utopia: no money, no coercion, no scarcity. The paradox Boucher and McAvan identify is that self-described libertarians and market fundamentalists have appropriated this left-wing imaginative vocabulary as a legitimating framework for AI development that is structurally the opposite of what Banks described.

The mechanism is what they analyze through the lens of the Culture’s intelligence agency, Special Circumstances: an organization that performs morally questionable operations in the present—covert manipulation, interference, occasional violence—on the grounds that these interventions will produce better outcomes for a larger population across a longer time horizon. Boucher and McAvan read Special Circumstances as a Schmittian state of exception built into

the utopian fabric: the sovereign decision to suspend normal ethical constraints is justified by appeal to the imagined future that the exception makes possible. Contemporary AI ideology, they argue, replicates this structure. The “necessary disruption,” the “transition cost,” the tolerance of present harm in service of the post-AGI abundance—all of these are Special Circumstances reasoning applied to actual economies and actual populations.

Future-Projection Capture

Within the formal framework of this essay, this ideological operation has a precise structure that extends the projection analysis to a temporal dimension.

Define a future projection operator π_f that maps the present trajectory space X_{present} onto an imagined future manifold M_{utopian} :

$$\pi_f : X_{\text{present}} \rightarrow M_{\text{utopian}}.$$

Longtermist utility aggregates value across time with weights w_t :

$$U = \sum_{t=0}^{\infty} w_t V_t.$$

When distant future value dominates—when $\lim_{t \rightarrow \infty} w_t V_t \gg \Phi_0$, where Φ_0 denotes present field constraints—present constraints become negligible in the objective.

Theorem 7 (Future Supremacy). *If longtermist utility weights satisfy $\lim_{t \rightarrow \infty} w_t V_t \gg \Phi_0$, then $\frac{\partial U}{\partial \Phi_0} \rightarrow 0$: present constraints become informationally invisible to the optimizing agent.*

Proof sketch. The gradient of U with respect to present constraint parameters Φ_0 is dominated by the far-future terms whenever their weight is sufficiently large. In the limit, Φ_0 drops out of the effective objective entirely. The institution or agent becomes unable to perceive present harm as a relevant cost; it is swamped by anticipated future benefit. This gives a formal derivation of why extreme longtermist frameworks can rationalize severe present harms: the future projection dominates the current field. \square

The structural isomorphism with wellness ideology is exact. Wellness ideology

says:

$$\rho_{\text{systemic}} \mapsto \rho_{\text{personal}}.$$

Techno-utopian AI ideology says:

$$\Phi_{\text{presentharm}} \mapsto \Phi_{\text{necessarytransitioncost}}.$$

In both cases, a real constraint is relabeled so that it becomes invisible to the operative decision process. The difference is only scale: wellness ideology captures individual reconstruction operators; techno-utopian ideology captures civilizational ones.

Boucher and McAvan connect this to what they characterize as TESCREAL—the cluster of transhumanism, extropianism, singularitarianism, cosmism, rationalism, effective altruism, and longtermism—and to the accelerationist strand associated with Nick Land, which explicitly celebrates the intensification of present disruption as a mechanism for reaching the post-human threshold faster. The Culture novels function as the imaginative legitimating manifold for all of these: a future of abundance is projected backward to justify present dislocation, and the imagined future population—including hypothetical digital beings of enormous scale—is used to make actually existing humans comparatively negligible in the utilitarian calculus.

Special Circumstances as Corporate State of Exception

The Schmittian framing is particularly useful. Carl Schmitt argued that the sovereign is whoever decides the exception: the moment of emergency that suspends normal legal and ethical constraints. Special Circumstances in Banks's novels is precisely this mechanism—presented sympathetically, as the price of long-run civilizational health.

What Boucher and McAvan show is that contemporary AI companies have internalized this structure without the democratic accountability that might justify it. The exception is declared unilaterally by a small number of technically capable actors who have decided, on behalf of humanity and future digital beings, that normal constraints—labor protections, regulatory oversight, democratic deliberation, precautionary principle—must be suspended for the duration of the transition.

In RSVP terms, this is Theorem 5 (Future Supremacy) applied institutionally: the

future projection π_f so dominates the objective that present field constraints Φ_0 become invisible. The map—the Culture, post-AGI abundance, the longtermist utility sum—eats the present territory, which continues to consist of actual people whose S_{future} is being compressed.

Regulatory Optionality: Law as Accessibility Preservation

Against the Race Frame

Tokson and Arbel [6] approach the same crisis from the legal side, and their central contribution maps directly onto the RSVP framework: adaptive regulation is not primarily a control mechanism but a method for preventing premature trajectory collapse.

Their starting point is the AI “arms race” or “race” metaphor that dominates both industry and policy discourse. The race frame asserts that because other actors—rival companies, rival nations—are moving as fast as possible toward advanced AI, any individual actor who slows down for safety or democratic deliberation will simply be overtaken, producing a worse outcome than if they had raced and won. Tokson and Arbel argue explicitly that this metaphor is pernicious because it structures the decision problem in a way that systematically removes safety-preserving choices. In the race frame, the only admissible action is acceleration; every other option is presented as equivalent to unilateral disarmament.

This is, in the formal vocabulary of this essay, a projection onto a one-dimensional manifold. The race frame reduces the full space of policy possibilities—a high-dimensional X_{policy} containing democratic deliberation, adaptive governance, international coordination, differentiated safety requirements, and precautionary pauses—to a single axis: faster versus slower. Theorem 3 (Projection Dominance) applies directly: as institutional dependence on this one-dimensional projection approaches unity, the decision process becomes blind to all information in the kernel, which contains most of the actual options.

Adaptive Regulation as Optionality Preservation

Tokson and Arbel’s positive proposal is adaptive regulation: governance frameworks designed not for a fixed anticipated future but for a range of possible futures, preserving the capacity to respond as conditions change. Their specific

mechanisms include legislative triggers (regulation that activates automatically when specified capability thresholds are crossed), disclosure requirements (forcing the production of information that would otherwise remain inside corporate kernels), critical-system oversight, and sunset and sunrise clauses that require active reauthorization rather than passive continuation.

Each of these is, in RSVP terms, a mechanism for maintaining $\mu(\mathcal{A}_t)$ —the measure of admissible future trajectories—against the pressure of accelerationist closure.

Theorem 8 (Regulatory Optionality). *Let \mathcal{A}_t be the set of admissible societal future trajectories at time t . Define $S_t = \log \mu(\mathcal{A}_t)$. Accelerationist policy reduces democratic branching: $\mu(\mathcal{A}_t) \downarrow$. Adaptive regulation preserves or increases branching: $\frac{d}{dt}\mu(\mathcal{A}_t) \geq 0$.*

Proof sketch. Accelerationist policy eliminates option value by committing irrevocably to a single development trajectory, closing off alternatives that would have remained accessible under more cautious governance. Each eliminated alternative reduces $|\mathcal{A}_t|$ and hence S_t . Adaptive regulation preserves option value by maintaining the legal and institutional infrastructure through which alternatives remain actionable: if conditions change, the governance framework can change with them. Tokson and Arbel’s specific mechanisms—triggers, disclosures, sunset clauses—are precisely the instruments that keep $\frac{d}{dt}\mu(\mathcal{A}_t) \geq 0$ by ensuring that no single development path is allowed to foreclose all others. \square

Precaution Under Irreversibility

A central argument in Tokson and Arbel is that policymakers should not wait for certainty before acting, because certain AI-related harms may be irreversible. This is a direct application of the accessibility entropy framework: once S_{future} collapses below a threshold, the trajectories that were eliminated cannot be recovered. The standard argument for regulatory delay—“wait until we know more”—ignores the asymmetry between reversible and irreversible harms. If the harm is reversible, delay costs the price of acting too late. If the harm is irreversible, delay costs the trajectories that could have been preserved, which are gone permanently.

In the RSVP framework, this asymmetry is formalized by the monotonicity of S_{future} under trajectory collapse: once $\mu(\mathcal{A}_t) = 0$ for some subset of formerly ad-

missible futures, those futures do not re-enter \mathcal{A}_t spontaneously. Precautionary regulation is the governance equivalent of brutal realism: it insists on keeping the residual ρ visible—the real distribution of possible futures—rather than allowing the accelerationist projection to substitute a single imagined future for the full range of what is actually possible.

The Projection Framework

Within the theoretical program of RSVP (Relativistic Scalar-Vector Plenum), a person inhabits a high-dimensional trajectory space X . Social and economic life requires the continuous production of compressed representations—projection maps π_s and π_e as defined above—whose target manifolds M_s and M_e capture the dimensions of a person that specific institutions are equipped to measure and reward.

Different institutions define different projections. The employer’s projection selects for reliability, productivity, and compliance signals. The social media platform’s projection selects for engagement, reach, and emotional resonance. The wellness industry’s projection selects for aspirational improvement trajectories. None of these projections is the person; all of them are lossy compressions.

Goffman’s dramaturgical model and Fromm’s market personality can now be given precise content within this framework. Goffman describes the process by which persons manage the production of contextually appropriate projections—the skilled manipulation of $\pi_s(x)$ for shifting audiences. Fromm describes the pathological limit in which the optimization pressure propagates backwards from M_s and M_e to X itself: the person reshapes their actual trajectory to conform to what the manifolds require, not merely the output they present. Theorem 1 above gives this the status of a dynamical result: when projection admissibility dominates the loss, gradient descent on \mathcal{L} necessarily produces this backwards propagation.

Trajectory Collapse: Theorem 3

The RSVP framework introduces accessibility entropy S as the log-measure of the volume of future trajectories accessible from a given state. Define the set of accessible futures for agent x at time t as

$$\mathcal{A}(x, t) = \{x(t + \tau) : x(t) \text{ remains viable under constraints}\},$$

and the RSVP accessibility entropy as

$$S_{\text{future}}(x, t) = \log \mu(\mathcal{A}(x, t)).$$

Theorem 9 (Trajectory Collapse). *If economic projection replacement removes income, institutional recognition, status, or bargaining power from agent x , then $S_{\text{future}}(x, t)$ decreases. Hence $\frac{d}{dt} S_{\text{future}} < 0$.*

Proof sketch. Each of income, status, and bargaining power is a constraint resource that maintains the viability of future trajectories. If $\pi_e(x)$ is automated and the institution withdraws the material and social goods previously conditional on $\pi_e(x)$, the set $\mathcal{A}(x, t)$ shrinks: formerly accessible futures (housing, healthcare, participation, political voice) require constraint resources that are no longer replenished. Hence $\mu(\mathcal{A}(x, t))$ decreases and S_{future} declines. \square

This is the formal meaning of mass “obsolescence” in the AI discussion’s sense. Humans do not disappear; their admissible future volume collapses. The transcript’s worry about disempowerment, welfare failure, and repression as responses to labor displacement all represent mechanisms through which S_{future} is further compressed once $\pi_e(x)$ has been automated away.

Brutal Realism as Residual Preservation: Theorem 4

Within the CLIO (Constraint-Limited Inference and Observation) framework, cognition performs a two-stage operation: projection followed by reconstruction. Let $R : M \rightarrow X'$ be the reconstruction map, so ordinary coping performs

$$X \xrightarrow{\pi} M \xrightarrow{R} X'.$$

Define the residual error as

$$\rho(x) = x - R(\pi x).$$

The residual is the information lost or distorted by the compression: the structural causes that do not appear in the projection, and therefore cannot appear in the reconstruction.

Theorem 10 (Brutal Realism as Residual Preservation). *Wellness ideology mini-*

mizes felt residual by reinterpreting ρ as personal defect:

$$\rho_{\text{systemic}} \mapsto \rho_{\text{personal}}.$$

Brutal realism suspends this false reconstruction by preserving ρ . Formally, it imposes the corrected reconstruction

$$R_{\text{BR}}(\pi x) = \pi x + \rho(x),$$

rather than replacing ρ with a socially acceptable explanation.

Proof sketch. If the reconstruction R is chosen to minimize social inadmissibility of the output—that is, if R is fitted to \mathcal{L}_s rather than to the true inverse of π —then $R(\pi x)$ will systematically attribute residual mismatch to variables within M_s (personal attitude, mindset, optimization level) even when the actual source lies outside M_s in the full trajectory space. R_{BR} corrects this by preserving ρ in the output, keeping the structural residual visible rather than absorbed into a false personal attribution. This is why the transcript frames brutal honesty as *accurate perception* rather than pessimism: it is the epistemically correct reconstruction, not an affectively negative one. \square

The CLIO interpretation of brutal realism is therefore a constraint on the reconstruction operator: do not fit R to the admissibility loss \mathcal{L}_s . Fit it to the field. Normal cognition, under projection supremacy, performs the former. Brutal realism is the discipline of performing the latter.

The Accessibility-Projection Conjecture

The five preceding theorems suggest a unifying quantitative relationship between projection dependence and accessible futures. Define the projection dependence ratio:

$$P = \frac{\lambda_s + \lambda_e}{\lambda_\Phi},$$

where λ_s is social projection pressure, λ_e is economic projection pressure, and λ_Φ is field constraint coupling.

Conjecture 2 (Accessibility-Projection).

$$\frac{dS_t}{dP} < 0.$$

Increasing dependence on projection systems strictly decreases accessible futures.

This conjecture unifies all four levels of the analysis. At the individual level (the Functional Melancholic): audience projections reduce accessible futures when $\lambda_s \gg \lambda_\Phi$. At the economic level (Lovely): economic projections reduce accessible futures when $\pi_e(x)$ is automated and income-maintaining trajectories are closed. At the ideological level (Boucher and McAvan): future-utopian projections reduce accessible futures when longtermist weights cause Φ_0 to vanish from the effective objective. At the legal level (Tokson and Arbel): regulatory foreclosure reduces accessible futures when the race frame collapses \mathcal{A}_t to a single track.

In each case, the mechanism is the same: a projection displaces field contact, and the information lost in $\ker(\pi)$ corresponds to trajectories that were admissible under the full field but become invisible—and then inaccessible—under the compressed representation.

An Accessibility Functional

The Accessibility-Projection Conjecture suggests a deeper variational structure. Define an accessibility functional:

$$\mathcal{F} = S_t - \alpha P + \beta C,$$

where $S_t = \log \mu(\mathcal{A}_t)$ is accessibility entropy, $P = \lambda_s + \lambda_e + \lambda_f$ is total projection pressure across social, economic, and future-ideological dimensions, $C = \lambda_\Phi$ is field constraint coupling, and $\alpha, \beta > 0$ are weighting parameters. The functional \mathcal{F} is offered as an organizing quantity rather than a derived physical entity: it is not a thermodynamic free energy but a scalar summary of the tradeoffs between accessibility preservation, projection pressure, and field coupling.

Theorem 11 (Accessibility Preservation Principle). *Healthy institutions evolve toward maximizing \mathcal{F} . Pathological institutions evolve toward maximizing P alone. Whenever projection optimization dominates field coupling,*

$$\frac{dS_t}{dt} < 0.$$

Proof sketch. If the institution maximizes \mathcal{F} , it jointly rewards high accessibility entropy (many admissible futures preserved), low projection pressure (field

contact maintained), and high field coupling (real constraints actively tracked). If instead the institution maximizes P alone—as accelerationist ideology and wellness optimization both do—field coupling is effectively set to zero, and by the Accessibility-Projection Conjecture, $dS_t/dt < 0$ follows directly. The result unifies individual burnout, AI labor replacement, longtermist future capture, and regulatory foreclosure as four components of \mathcal{P} in which P dominates C . \square

The functional \mathcal{F} is the central object of the RSVP/CLIO theory of institutional health. It measures not whether a society is productive or efficient in any given projection but whether it is preserving the conditions under which meaningful future choice remains possible.

The Societal Case and Full Synthesis

The individual pathology described by Theorems 1 and 4—a single agent whose objective gradient has been rotated toward projection admissibility, and whose reconstruction operator suppresses the systemic residual—scales through three further levels: the economic (Theorems 2 and 3), the ideological (Theorem 5), and the legal (Theorem 6).

At the individual scale: the market personality and the performance of recovery. λ_s and λ_e dominate \mathcal{L}_Φ ; R maps systemic residuals to personal defects. At the economic scale: AI-driven labor projection replacement. Institutions optimize M_e while the underlying population of trajectories X loses accessible future volume. At the ideological scale: techno-utopian future-projection capture. Longtermist weights render Φ_0 invisible; present harm becomes transition cost. At the legal scale: accelerationist regulatory foreclosure. The race frame collapses \mathcal{A}_t to a single track, eliminating the democratic branching that adaptive governance would preserve.

All four instantiate the same underlying dynamic. The Projection Crisis is not an arithmetic sum of incommensurable terms but a decomposition into four simultaneous, mutually reinforcing components:

$$\mathcal{P} = (P_{\text{phen}}, P_{\text{econ}}, P_{\text{ideo}}, P_{\text{legal}})$$

The corrective at each level is recoupling—restoring the sensitivity of the opera-

tive objective gradient to real field constraints:

$$\text{epistemic health} = \frac{\|\nabla \mathcal{L}_\Phi\|}{\|\lambda_s \nabla \mathcal{L}_s + \lambda_e \nabla \mathcal{L}_e\|}$$

When this ratio rises, trajectory viability reasserts itself against projection management at every scale simultaneously.

Brutal realism, AI reform, critique of longtermist ideology, and adaptive regulation are all, in this formal sense, attempts to raise the ratio—to recouple the gradient of action to the field rather than to the map. They differ in their target but share the same structural aim. The deepest thesis is therefore:

Brutal realism restores epistemic contact with present constraints. AI reform restores democratic access to future economic trajectories. Critique of techno-utopian ideology restores visibility of present harm against future-projection capture. Adaptive regulation preserves optionality against accelerationist closure.

And the unifying governance principle, elevated from regulatory prescription to a general theory of civilizational resilience:

$$\text{Civilizational health} = \text{preservation of trajectory diversity}$$

Accessibility S_t is ultimately a measure of remaining trajectories. When it is preserved, the four domains of the analysis map onto four kinds of trajectory diversity that a healthy civilization maintains simultaneously:

Brutal realism preserves *cognitive trajectory diversity*: the capacity of individuals to hold accurate models of their actual constraints rather than admissibility-optimized projections of them. AI reform preserves *economic trajectory diversity*: the range of viable material futures accessible to populations whose labor projections are being automated. Democratic critique of techno-utopian ideology preserves *political trajectory diversity*: the openness of the future against the closure imposed by any single civilizational narrative, however compelling. Adaptive regulation preserves *institutional trajectory diversity*: the capacity of governance to respond to what actually emerges rather than to what any projection predicted.

This formulation elevates the analysis from a critique of AI and wellness culture

into a general theory of how civilizations lose contact with reality through projection capture and how they recover through accessibility preservation. The paper’s claim is not that any particular future is correct, but that the institutional conditions for choosing among futures must themselves be preserved—that \mathcal{F} must be maximized, that $\mu(\mathcal{A}_t)$ must not be collapsed prematurely, and that the map must be prevented from eating the territory while the territory is still where we live.

The Projection Crisis: Implications

The four theorems established above jointly characterize the projection crisis. Theorem 1 shows how agents are dynamically driven away from trajectory viability toward projection admissibility when social and economic loss terms dominate. Theorem 2 shows how economic substitution targets projections rather than persons, making reskilling an inadequate response. Theorem 3 shows how projection replacement produces measurable collapse of accessible future volume, formalizing “obsolescence” as a decline in S_{future} rather than metaphysical displacement. Theorem 4 shows how wellness ideology compounds this by misdirecting the reconstruction operator—misattributing systemic residuals as personal defects—and how brutal realism corrects this by preserving ρ in the output.

Contemporary society exhibits projection supremacy across every institutional dimension simultaneously: social media platforms value engagement projections, employers value employability projections, markets value consumer-preference projections, universities value credential projections, and AI systems are trained on projections of all these kinds. The underlying trajectory spaces—the full dimensionality of human lives, tacit knowledge, embodied skill, and temporal development—become progressively invisible to the systems that most consequentially shape the conditions of existence.

The standard debate about AI employment effects focuses on job counts. That debate is conducted almost entirely within M_e . The deeper question is whether the persons whose projections are automated retain access to a sufficient volume of $\mathcal{A}(x, t)$. Jobs are one mechanism by which that access is maintained; they are not the ultimate object of concern, and policies that update $\pi_e(x)$ without expanding S_{future} do not address the underlying problem.

Institutional Design as the Decisive Variable

Neither the cultural analysis nor the economic analysis supports technological determinism. The same capability that enables projection extraction at scale also enables new forms of abundance, coordination, and discovery. The technical capability does not determine the outcome; the institutional context does.

Compression is ubiquitous and necessary; the pathology arises when the compression is optimized for a single class of values at the expense of all others, and when the resulting pressure reshapes the trajectory spaces being compressed. Institutions oriented toward trajectory preservation would ask not only “what projection does this person currently offer?” but “what volume of accessible futures does this person require, and what conditions support the expansion of that volume?”

That reorientation is a political and institutional challenge, not a technical one. Brutal realism, AI reform, and trajectory-preserving institutional design are all formal attempts to raise the epistemic health ratio established above—to recouple the gradient of action to the field rather than to the map. The first requirement for any such reorientation is accurate diagnosis: the refusal, formalized in Theorem 4, to translate the systemic residual ρ_{systemic} into the personal residual ρ_{personal} . Naming conditions accurately does not solve them, but it is the precondition for interventions that address causes rather than symptoms—and, as the first transcript notes, clarity is less exhausting than fantasy, even when what becomes clear is not good [2].

Reality, Models, and Recoupling

The argument developed throughout this essay should not be interpreted as a rejection of abstraction, modeling, measurement, or institutional coordination.

Human beings encounter reality through representations at every scale. Scientific theories, legal categories, economic indicators, linguistic concepts, and personal narratives all function as projection systems that render complexity tractable. The problem identified here is therefore not abstraction itself but abstraction that loses awareness of its own limitations.

Recoupling is the process through which a projection remains corrigible in the presence of information originating outside its representational boundaries. A healthy institution is not one that avoids projection. It is one that maintains

mechanisms through which information from the projection kernel $\ker(\pi)$ can re-enter decision-making processes: feedback channels, residual monitoring, qualitative observation, democratic input, and the tolerance of results that do not fit the model.

The central danger of projection supremacy is not that institutions use models. It is that institutions increasingly mistake models for the domains from which those models were generated. A civilization remains adaptive to the extent that it preserves pathways through which territories can continue correcting maps—through which the residual $\rho(x) = x - R(\pi x)$ remains visible rather than absorbed into a false reconstruction.

Conclusion: The Question of What Society Is For

The AI discussion ultimately raises a question that transcends the technology: what is a society for?

If productivity continues increasing while human labor becomes less necessary to the operation of the dominant economic projections, then the traditional equation

$$\text{income} = \text{employment}$$

becomes structurally unstable. The instability is not new; it has been building throughout the history of industrial automation. AI accelerates it and extends it to domains previously assumed to be insulated.

The question of what a society is for becomes unavoidable when productive capacity is no longer tightly coupled to human labor. It is a question about the distribution of accessible futures. If projection supremacy is the dominant mode of social organization, then the answer given by existing institutions will be: a society is for the optimization of economically valuable projections. The persons behind the projections will be valued to the extent that their projections remain economically legible, and will find their accessible future volumes compressed to the extent that their projections become automatable.

That is not a satisfactory answer. But identifying why it is unsatisfactory requires the kind of analysis developed here. The analysis of brutal realism identifies the category error: structural conditions misattributed to individual deficiencies. The analysis of projection extraction identifies the mechanism: the industri-

alization of economically legible manifolds at the expense of the trajectories they represent. The analysis within RSVP and CLIO identifies the formal consequence: the compression of S_{future} for populations whose trajectory spaces are no longer valued by the dominant projection functions.

And the synthesis identifies the path forward—not as a specific policy or technical solution, but as an epistemic and institutional reorientation. Institutions capable of resisting projection supremacy are institutions that maintain explicit awareness of the gap between the map and the territory: between the compressed manifold they optimize and the full trajectory space of the persons whose lives they shape.

Brutal realism is the name for the epistemic practice of keeping that gap visible. It is, in the end, not about burnout, or wellness culture, or artificial intelligence. It is about the conditions under which a civilization can remain capable of recognizing value that is not immediately reducible to productivity, prediction, optimization, or economic output—and about the cost of losing that capacity.

Author: Flyxion. Independent researcher. <https://github.com/standardgalactic>.

References

- [1] Jean-Pierre Aubin. *Viability Theory*. Birkhäuser, Boston, 1991.
- [2] The Functional Melancholic [@functionalmelancholic]. *The Strange Comfort of Brutal Realism*. YouTube, 2025.
- [3] Novara Media [@NovaraMedia]. *Society Is About To Change. And No One Is Ready | Richard Hames meets Garrison Lovely*. YouTube, 2025.
- [4] Garrison Lovely. *Obsolete: The AI Industry's Trillion-Dollar Race to Replace You and How to Stop It*. 2025.
- [5] Geoff M. Boucher and Emily McAvan. The Digital Tech Brologarchy's Interest in Left-Wing Science Fiction: A Critical Reading of the Culture of Techno-Libertarianism. *tripleC: Communication, Capitalism & Critique*, 24(1):1–22, 2026.
- [6] Matthew Tokson and Yonathan Arbel. Artificial Intelligence and Existential Risk. *Connecticut Law Review*, 59, 2027 (forthcoming). <https://ssrn.com/abstract=6288138>