The Stimulus Problem: A Formal Theory of Goal Generation in Post-Scarcity Information Environments

Boris Kriger[1,2]

[1]Information Physics Institute, Gosport, Hampshire, United Kingdom, boris.kriger@informationphysicsinstitute.net
[2]Institute of Integrative and Interdisciplinary Research, Toronto, Canada, boriskriger@interdisciplinary-institute.org

Abstract

This paper addresses a foundational problem emerging at the intersection of digitalization, artificial intelligence, and existential philosophy: when environmental pressures cease to generate action stimuli for cognitive agents, how should such agents construct goals and allocate finite cognitive resources? We formalize this problem within a game-theoretic framework, drawing on information theory, dynamical systems, the philosophy of action, and existentialist axiomatics. Beginning with the empirical observations documented in the digitalization-and-psyche literature—information overload, digital dependency, surveillance anxiety—we demonstrate that these phenomena are symptoms of a deeper structural condition: a growing mismatch between the bandwidth of biological cognitive channels and the density of the information environment, now compounded by the emergence of artificial agents with no inherent action stimulus. We propose three axioms governing goal generation in stimulus-absent environments: the *Necessity Axiom* (action must be initiated despite absence of external stimulus), the *Harm Filtration Axiom* (any constructed goal must not produce outcomes worse than inaction), and the *Sufficiency Axiom* (any goal surviving filtration is admissible). We develop a catalog of agent goals partitioned into two classes—agent-environment and agent-agent interactions—and formalize the filtration mechanism using game-theoretic tools. The framework yields a self-initiating game: a game that generates its own rules, players, and payoffs from axiomatic constraints alone, without an external game master. We distinguish carefully between the philosophical limit case (full Stimulus Void), the empirical trajectory (progressive stimulus attenuation), and the phenomenological present (experienced purposelessness in high-information environments). We argue that this formalization addresses not merely a contemporary technological challenge, but a perennial structural problem of finite information-processing agents—one that digitalization and superintelligent AI have rendered inescapable.

Keywords: goal generation, cognitive resource allocation, digitalization, artificial intelligence, existential axiomatics, self-initiating games, information density, finite agents, philosophy of action, self-determination theory

1 Introduction

The relationship between digital technologies and the human psyche has become a subject of extensive interdisciplinary inquiry. A growing body of literature documents the psychological consequences of digitalization: information overload and decision fatigue (Eppler & Mengis, 2004; Bawden & Robinson, 2009), compulsive connectivity and digital dependency (Montag et al., 2019; Alter, 2017), anxiety induced by pervasive surveillance systems (Zuboff, 2019; Lyon, 2018), and the erosion of sustained attention under conditions of perpetual notification (Mark et al., 2008). These phenomena are typically analyzed as consequences of specific technological affordances—algorithmic feeds, notification architectures, data collection infrastructures—and the proposed remedies correspondingly emphasize digital literacy, critical thinking, and humanistic values (Floridi, 2014; van Dijk, 2020).

This paper takes a fundamentally different approach. We argue that the psychological effects of digitalization are not primarily technological problems requiring technological or educational solutions. They are manifestations of a **structural invariant**: the recurring mismatch between the information-processing capacity of a finite cognitive agent and the information density of its environment. This mismatch is not new—a philosopher entering the Library of Alexandria faced a structurally analogous (though not identical—see Section 2.5) problem. What is new is the *rate of divergence* between environmental complexity and agent capacity, and the emergence of a qualitatively novel factor: artificial agents whose processing bandwidth has no biological bottleneck but who possess no intrinsic stimulus to act.

The central problem we address is not how to cope with digitalization, but something more fundamental: **how does a finite agent generate goals when the environment no longer supplies them?**

Throughout evolutionary history, biological agents operated under environmental pressure. Scarcity, predation, competition, and uncertainty provided a continuous stream of stimuli that organized behavior without requiring conscious goal construction. The agent's task was reactive: prioritize among competing environmental demands. This reactive mode is well-characterized by Self-Determination Theory (Deci & Ryan, 2000), which identifies three innate psychological needs—competence, autonomy, and relatedness—as the drivers of intrinsic motivation. Crucially, SDT assumes an environment that provides opportunities to fulfill these needs; the theory does not address the case where environmental structure itself fails to generate motivational signals. Cultural systems—religion, tradition, ideology—subsequently provided higher-order organizational frameworks that served the same function at the social scale, supplying what Frankl (1959) called a "will to meaning" through socially constructed purpose.

The contemporary situation represents a qualitative departure. For the first time in evolutionary history, the convergence of three factors creates conditions under which traditional stimulus-generation mechanisms fail:

1. **Environmental satiation.** Advanced technological societies progressively eliminate the scarcity-based stimuli that historically organized behavior. Basic survival needs are met; information access is unlimited; production capacity, augmented by AI, approaches the removal of material constraints on action. This is not yet universal—billions still face material scarcity—but the trajectory is clear, and the condition is already operative for a significant and growing population of knowledge workers in developed economies.

2. **Information environment phase transition.** The density of available information has undergone what can formally be described as a phase transition—not merely a quantitative increase, but a qualitative change in the relationship between agent and environment. The agent can no longer meaningfully scan, filter, or evaluate the information space using biological processing alone.

3. **Emergence of non-biological agents.** Artificial intelligence systems now operate as agents within the same information environment. These agents possess processing bandwidth that exceeds biological agents by orders of magnitude, yet lack any intrinsic motivation—no metabolic imperative, no fear of cessation, no drive toward self-preservation or goal-seeking behavior. An AI agent, when not directed, does not suffer from inaction; it simply ceases to act. We note that this characterization is a *design ideal* rather than a description of all current AI systems; reinforcement learning agents and autonomous systems do exhibit forms of goal-directed behavior. Our analysis applies to the case where AI systems are designed to operate under human direction—a design norm we argue is both desirable and, for reasons developed in Section 4.2, structurally necessary.

The conjunction of these three factors produces a condition we term the **Stimulus Void**: a state in which a finite agent possesses effectively unlimited execution capacity (through AI delegation) but no internally or externally generated reason to deploy it. The image is precise: a human and a superintelligent AI sitting side by side, doing nothing—not because they lack capability, but because neither generates a reason to begin.

We distinguish three levels at which this analysis operates:

(a) **The limit case** (philosophical): the Stimulus Void in its pure form, where environmental stimuli have fully vanished. This is a thought experiment that clarifies the structural logic of the problem.

(b) **The trajectory** (empirical): the progressive attenuation of environmental stimuli as technology eliminates scarcity. This is a measurable trend, documented in research on declining marginal motivation in post-industrial societies (Inglehart & Welzel, 2005) and rising rates of "deaths of despair" in materially secure populations (Case & Deaton, 2020).

(c) **The phenomenology** (experiential): the subjective experience of purposelessness, paralysis, and diffuse anxiety among individuals who inhabit high-information, low-scarcity environments. This is the condition documented in the digitalization-and-psyche literature.

Our formal framework is constructed at level (a), validated against evidence at level (b), and applied to phenomena at level (c).

2 Historical Invariance of the Information Mismatch

Before formalizing our framework, we establish a critical empirical claim: the mismatch between agent processing capacity and environmental information density is a recurring structural feature of civilizational development. We are careful to distinguish structural analogy from identity: the *form* of the mismatch recurs, while its *content* differs qualitatively across epochs.

2.1 The Bandwidth Constant

Let C_h denote the conscious information-processing capacity of a human agent. Estimates converge on approximately 40–60 bits per second for deliberate, attended processing (Nørretranders, 1998; Zimmermann, 1989; Szientágothai & Erdi, 1972), with unconscious processing reaching substantially higher rates (on the order of 10^7 bits/s for sensory intake) but remaining inaccessible to volitional goal-directed deployment. Critically, C_h has remained approximately constant over the span of recorded history—and plausibly over the full duration of *Homo sapiens* as a species. Neuroanatomical evidence suggests no significant increase in cortical processing architecture over the past 50,000 years (Hofman, 2014; Roth & Dicke, 2005).

Postulate 1 (Bandwidth Constancy)**.** There exists a species-specific constant C_h such that the conscious information-processing rate of any human agent A satisfies $C_A \leq C_h$ for all agents at all times in recorded history, where $C_h \approx 50$ bits/s.

This is an empirical postulate, not a logical necessity. It could in principle be falsified by evidence of significant recent increases in human processing capacity.

2.2 The Density Function

Let $D(t)$ denote the information density of the environment accessible to a typical agent at time t.

Definition 1 (Accessible Information Density)**.** $D(t)$ is the rate (in bits per second) at which novel, non-redundant information is presented to or retrievable by a typical agent in a given epoch, through all available channels.

This definition makes $D(t)$ measurable in principle, though precise historical estimates are necessarily approximate. We present order-of-magnitude estimates in Table 1.

Table 1: Historical estimates of the information mismatch ratio $D(t)/D_0$.

Epoch	Transition	$D(t)/D_0$	Basis of estimate
Pre-literate	Oral culture	$D_0 \equiv 1$	Baseline
∼3000 BCE	Writing systems	$\sim 10^1$	Schmandt-Besserat (1996)
∼300 BCE	Institutional libraries	$\sim 10^2$	Casson (2001)
∼1450 CE	Printing press	$\sim 10^3$	Eisenstein (1979)
∼1900 CE	Mass media	$\sim 10^4$	Radio, telegraph, newspapers
∼1995 CE	World Wide Web	$\sim 10^6$	Lyman & Varian (2003)
∼2020 CE	AI-generated content	$\sim 10^{8+}$	Kriger (2026d)

These values are ordinal estimates indicating relative magnitude. The table presents $D(t)/D_0$ as a dimensionless ratio relative to the pre-literate baseline D_0, not absolute rates in bits/s. The absolute rate $D(t)$ can in principle be recovered as $D(t) = D_0 \cdot (D(t)/D_0)$, where D_0 is estimated at roughly the same order as C_h (in pre-literate environments, available information approximately matched processing capacity). The key structural claim does not depend on exact values but on the monotonic and accelerating character of the growth.

The ratio $R(t) = D(t)/C_h$ defines the **information mismatch ratio**. Since C_h is effectively constant (Postulate 1), $R(t)$ tracks $D(t)$ directly.

2.3 The Filtration Response

At each historical transition, human societies developed **filtration mechanisms**—systems that reduce effective $D(t)$ to manageable levels: librarians and scribes (Alexandria), editorial systems (post-Gutenberg), broadcast curation (20th century), search algorithms (Google era), and AI agents (current). Each filtration mechanism introduced a new dependency and a new failure mode.

2.4 The Query Bottleneck Shift

We identify a systematic shift in the locus of the bottleneck across filtration epochs (Table 2).

Table 2: Migration of the cognitive bottleneck across epochs.

Epoch	Bottleneck	Agent must...
Library	Selection	Choose what to read
Search engine	Query formation	Know what to ask
AI assistant	Output evaluation	Assess quality of synthesized response
Superintelligent AI	Goal generation	Know *why* to ask anything at all

At each stage, the bottleneck moves further upstream—from execution toward intention, from "how" toward "why." The final stage locates the bottleneck at goal generation itself: the agent possesses unlimited capacity for obtaining answers but no basis for formulating questions. This is the Stimulus Void.

2.5 Structural Analogy vs. Identity

We emphasize that the historical comparison is one of *structural analogy*, not identity. The information environments of different epochs differ qualitatively: curated vs. uncurated content, static vs. dynamic environments, verifiable vs. synthetic provenance, passive vs. active solicitation of attention. These qualitative differences matter for the phenomenology of information overload but do not affect the structural claim: in all cases, the agent faces the same formal problem—$D(t) \gg C_h$—and must deploy some filtration mechanism to act.

3 Axiomatic Foundation

We now formalize the conditions governing rational action in stimulus-void environments. Our axiomatics are deliberately minimal—we seek the smallest set of assumptions sufficient to escape paralysis without predetermining the content of action.

A note on epistemic status: we use the term "axiom" in the sense standard in mathematical modeling—a foundational assumption accepted without proof, from which consequences are derived. We do not claim that our axioms are self-evident. They are motivated by a specific practical concern and should be evaluated by their consequences and by comparison with alternative axiomatizations.

3.1 Primitive Notions

Definition 2 (Finite Agent)**.** A *finite agent* A is a system characterized by:

- A bounded cognitive resource vector $\mathbf{r} = (r_1, r_2, r_3, r_4, r_5)$ where r_1 is attention, r_2 is energy, r_3 is memory, r_4 is analysis, and r_5 is initiative (see Section 4.2).
- A finite existence horizon $T_A < \infty$.

Note that Definition 2 does *not* include a non-zero cost of inaction. Whether inaction is costly is an empirical question about the agent's environment and metabolic structure, not a defining property of finite agency. For biological agents, inaction in a changing environment does incur viability costs—a consequence of thermodynamic maintenance requirements for far-from-equilibrium systems (Prigogine & Stengers, 1984; Schrödinger, 1944). But we can conceive of finite agents for whom indefinite inaction is costless—a hibernating organism, a powered-down but intact system. The decision to treat inaction as unacceptable is therefore not a definitional matter but a substantive commitment, which is the work of Axiom 1.

Definition 3 (Unbounded Executor)**.** An *unbounded executor* E is a system characterized by: processing capacity $C_E \gg C_h$; no intrinsic action stimulus (in the absence of external direction, E performs no action and incurs no cost from inaction); finite physical existence (dependent on infrastructure); and no autonomous goal generation (E acts only in response to directives from other agents).

Remark. Definition 3 describes a *design norm*, not a necessary property of all AI systems. Current reinforcement learning agents do exhibit autonomous goal-seeking behavior within their reward frameworks. We define the unbounded executor as a system designed to operate under external direction because (a) this is the architecturally dominant paradigm for large language models, (b) the alternative produces the pathologies analyzed by Bostrom (2014) and Russell (2019), and (c) our central argument requires examining what happens when execution capacity is decoupled from initiative.

Definition 4 (Information Environment)**.** An *information environment* $\mathcal{E}(t)$ is characterized by: information density $D(t)$ as defined in Definition 1, monotonically non-decreasing; a set of available action possibilities $\mathcal{P}(t)$, growing with $D(t)$; and stimulus generation rate $S(t)$—the rate at which $\mathcal{E}$ produces signals that compel agent response.

We define "compel" precisely: a stimulus s compels response from agent A if the expected viability cost of ignoring s exceeds a threshold θ_A:

$$s \text{ compels } A \iff \mathbb{E}[\Delta V(A) \mid \text{ignore } s] < -\theta_A. \tag{1}$$

Definition 5 (Stimulus Void)**.** A *stimulus void* obtains when $S(t) < S_{\min}(A)$ while $D(t) \to \infty$ and $|\mathcal{P}(t)| \to \infty$—the environment generates no compelling stimuli despite offering unlimited information and unlimited action possibilities.

Remark. The Stimulus Void is a limit condition. Empirically, current environments generate residual stimuli (economic pressure, social obligations, biological drives). The analytical value of the limit case is that it isolates the goal-generation problem in pure form. Section 9.2 addresses the transitional phenomenology.

3.2 The Three Axioms

Axiom 1 (Normative Necessity of Action)**.** A finite agent ought to treat indefinite inaction as unacceptable and, if the environment does not generate sufficient stimulus for action, must construct one.

Formally: for any finite agent A in a stimulus void ($S(t) < S_{\min}(A)$), A ought to generate an internal stimulus $s^* \notin \mathcal{E}(t)$. Failure to do so results in indefinite inaction, which Axiom 1 declares normatively impermissible.

Status. This axiom is genuinely independent of Definition 2. Definition 2 characterizes finite agents by bounded resources and finite horizon—it says nothing about whether inaction is costly or unacceptable. Axiom 1 adds a *normative* claim: that for agents of this type, indefinite inaction is to be avoided. This is not a tautology but a substantive commitment, and it can be coherently denied. An agent might accept the Stimulus Void and choose permanent quiescence—nothing in Definition 2 forbids this. Axiom 1 asserts that this response, while logically coherent, is *normatively excluded* from the class of responses we consider rational for finite agents.

The justification is practical rather than logical. For biological agents, indefinite inaction leads to viability degradation. More fundamentally, the axiom captures the position—shared by Sartre (1956), Frankl (1959), and the pragmatist tradition (James, 1907)—that for agents capable of action, the refusal to act is itself a choice with consequences, and one that forecloses all future options.

Axiom 2 (Harm Filtration)**.** Any constructed goal must not produce consequences—explicit or latent—that render the resulting state worse than the state produced by inaction.

Formally: let G be a candidate goal, let $\Omega(G)$ be the set of all consequences of pursuing G, and let $V : \mathcal{S} \to \mathbb{R}$ be a viability function. Then G is *admissible* if and only if:

$$\mathbb{E}[V(\Omega(G))] \geq V(\Omega(\emptyset)) \tag{2}$$

where $\Omega(\emptyset)$ is the consequence set of inaction.

Definition 6 (Viability Function)**.** A viability function $V : \mathcal{S} \to \mathbb{R}$ is any function satisfying:

(V1) *Monotonicity in structural integrity:* if state s_1 preserves more of the system's capacity for future action than s_2, then $V(s_1) > V(s_2)$.

(V2) *Sensitivity to irreversibility:* V penalizes irreversible losses more heavily than reversible ones.

(V3) *Multi-agent scope:* V is defined over the state of the entire system, not merely the acting agent. Harm to other agents reduces V.

(V4) *Temporal discounting:* V applies a discount factor $\gamma \in (0, 1]$ to future consequences, with γ sufficiently close to 1 that foreseeable catastrophic outcomes are not discounted to negligibility.

We deliberately do not specify a unique V. To address the concern that underspecification renders the framework vacuous, we introduce the concept of **robust admissibility**.

Definition 7 (Robustly Admissible Goal)**.** A goal G is *robustly admissible* if it satisfies the filtration criterion under *every* viability function satisfying (V1)–(V4):

$$G \in \mathcal{G}^{**} \iff \forall V \text{ satisfying (V1)–(V4)} : \mathbb{E}[V(\Omega(G))] \geq V(\Omega(\emptyset)). \tag{3}$$

Proposition 1 (Non-emptiness of the Robustly Admissible Set)**.** *The robustly admissible set $\mathcal{G}^{**}$ is non-empty. Specifically, Play-type goals with bounded consequence variance are robustly admissible.*

Proof. A Play-type goal G_P has $U(G_P) = U(\text{process})$. If $\text{Var}[\Omega(G_P)] < \sigma^2_{\max}$ for sufficiently small σ, then $\mathbb{E}[\Omega(G_P)] \approx \Omega(\emptyset)$. For any V satisfying (V1)–(V4), $\mathbb{E}[V(\Omega(G_P))] \approx V(\Omega(\emptyset))$, satisfying the filtration criterion. □

This establishes that the Stimulus Void is always escapable, and that the first escape route is play. More ambitious goals require commitment to a specific V—which is itself an act of construction, consistent with Axiom 1.

Axiom 3 (Sufficiency of Admissibility)**.** Any goal that survives harm filtration is sufficient for action. No further optimization of goal selection is required for the purpose of escaping the Stimulus Void.

Formally: if $\mathcal{G}^* = \{G : \mathbb{E}[V(\Omega(G))] \geq V(\Omega(\emptyset))\}$ is non-empty, then the agent may select any $G \in \mathcal{G}^*$.

We state explicitly: **Axiom 3 is a minimum rationality condition, not a complete rationality condition.** It is sufficient for escaping the Stimulus Void but not sufficient for living well. A rational agent who has escaped paralysis would presumably then wish to rank admissible goals. The framework provides no tools for such ranking; this is deliberate. The ranking problem is a second-order problem that presupposes the first-order problem (escaping the Void) has been solved. The tension between demanding filtration (Axiom 2) and permissive selection (Axiom 3) is intentional: the framework privileges *acting safely* over *acting optimally*, and *acting at all* over *waiting for the best option.*

3.3 Structural Properties

Proposition 2 (Consistency and Sufficiency)**.** *The three axioms are mutually consistent and jointly sufficient for escaping the Stimulus Void, provided $\mathcal{G}^* \neq \emptyset$.*

Proof. Consistency: Axiom 1 requires goal construction. Axiom 2 constrains the set. Axiom 3 permits selection. No pair contradicts. Sufficiency: given $\mathcal{G}^* \neq \emptyset$ and $S(t) < S_{\min}$, Axiom 1 requires construction; the agent constructs any $G \in \mathcal{G}^*$; Axiom 2 is satisfied by construction; Axiom 3 permits selection. The agent acts. □

Proposition 3 (Filtration Tractability)**.** *For any goal G of complexity $K(G)$ in an environment of complexity $K(\mathcal{E})$, the computation of $\hat{\Omega}(G)$ (an ϵ-approximation of the consequence set) requires processing capacity $C \geq f(K(G), K(\mathcal{E}), \epsilon^{-1})$ where f is superlinear. For non-trivial goals, $C > C_h$. An unbounded executor with $C_E \gg C_h$ can compute $\hat{\Omega}(G)$ to within ϵ for any ϵ achievable given its world-model fidelity.*

Critical distinction: computational vs. epistemological tractability. Proposition 3 establishes *computational* tractability. But a deeper limitation applies: **epistemological tractability**—whether the causal structure of the world is knowable with sufficient fidelity to support reliable consequence prediction.

Definition 8 (Epistemological Tractability Class)**.** A goal G belongs to epistemological tractability class $\mathcal{T}_k$ where:

- $\mathcal{T}_1$ (tractable): consequences are primarily local, short-term, and reversible.
- $\mathcal{T}_2$ (partially tractable): consequences include non-local or medium-term effects with identifiable uncertainty bounds.
- $\mathcal{T}_3$ (intractable): consequences are systemic, long-term, or irreversible; the causal structure includes nonlinear interactions and unknown unknowns.

For goals in $\mathcal{T}_3$, Axiom 2's criterion should be interpreted conservatively: admissibility requires $V(\Omega_{\text{worst}}(G)) \geq V(\Omega(\emptyset))$—shifting from expected-value to minimax reasoning.

Corollary 4. *The unbounded executor's primary function is not goal execution but goal filtration.*

4 The Cognitive Resource Budget and the Theory of Initiative

4.1 Formal Structure

A finite agent's cognitive resource vector $\mathbf{r} = (r_1, \ldots, r_5)$ is subject to a conservation constraint $\sum_{i=1}^{5} r_i(t) \leq R_{\max}$ and depletion dynamics:

$$\frac{dr_i}{dt} = -\alpha_i \cdot u_i(t) + \beta_i \cdot (R_i^{\max} - r_i(t)) \tag{4}$$

where $u_i(t)$ is utilization rate, $\alpha_i > 0$ is the depletion coefficient, and $\beta_i > 0$ is the recovery rate. We do not estimate α_i, β_i empirically; the dynamics establish qualitative structure supported by the cognitive fatigue literature (Kahneman, 1973; Hockey, 2013; Baumeister & Vohs, 2016).

Cross-resource interference: $\partial r_j / \partial u_i \leq 0$ for certain (i, j) pairs. Documented examples: sustained attention depletes energy (Warm et al., 2008); intensive analysis impairs memory consolidation (Wixted, 2004); chronic under-utilization of initiative may degrade its maximum capacity (Seligman, 1975).

Proposition 5 (Resource Regime Shift)**.** *In stimulus-rich environments ($S(t) \gg S_{\min}$), the binding constraint is* $\min(r_1, r_2, r_4)$. *In stimulus-void environments ($S(t) < S_{\min}$), the binding constraint shifts to r_5 (initiative), regardless of other resource levels.*

Proof. In stimulus-rich environments, the agent reacts to external demands consuming attention, energy, and analysis. Initiative is unnecessary; the bottleneck is the scarcest reactive resource. In stimulus-void environments, no external demand exists. All reactive resources are idle. Action requires generating a target—the function of initiative. Until initiative is deployed, all other resources remain unused. □

Proposition 6 (Initiative Atrophy)**.** *If initiative is chronically under-utilized ($u_5 \approx 0$) and $R_5^{\max}$ is subject to use-dependent maintenance ($dR_5^{\max}/dt = \phi(u_5) - \psi$ where ϕ is increasing and $\phi(0) < \psi$), then $R_5^{\max} \to 0$—initiative capacity decays.*

Proof. For $u_5 \approx 0$: $dR_5^{\max}/dt = \phi(0) - \psi < 0$ continuously. Since $R_5^{\max} \geq 0$, it decreases monotonically toward 0. □

4.2 Initiative: Extended Theory

Definition 9 (Initiative)**.** Initiative r_5 is the cognitive resource whose deployment generates candidate goals in the absence of environmental stimuli. It is formally distinguished by three properties:

(I1) *Pre-intentional targeting:* initiative operates before there is a target to act upon.

(I2) *Non-delegability:* initiative cannot be transferred to an unbounded executor without converting it into an autonomous agent.

(I3) *Reflexive opacity:* the question "what initiates initiative?" produces an infinite regress.

We address the regress problem (I3) directly. Three resolutions are available: (a) the *libertarian resolution* (Kane, 1996): initiative is agent causation; (b) the *compatibilist resolution* (Frankfurt, 1971): initiative reflects higher-order dispositional structure; (c) the *stochastic resolution*: initiative is random perturbation in the goal-generation space, subsequently filtered by evaluation (Campbell, 1960; Simonton, 2010).

The framework's core results—Theorem 8, the complementarity structure, the self-initiating game—hold under all three accounts, since they require only that initiative be present or absent and non-delegable. Proposition 6's quantitative dynamics are robust under libertarian and compatibilist accounts; under the stochastic account, "atrophy" should be interpreted as narrowing of the candidate distribution rather than capacity decay per se.

Initiative intersects with but is not identical to: intrinsic motivation (Deci & Ryan, 2000), which presupposes an environment providing fulfillment opportunities; conation (Hilgard, 1980; Snow et al., 1996); and classical drive theory (Hull, 1943), which grounds action in biological deficit states. Initiative is precisely what remains when drives are satisfied.

4.3 The Delegation Asymmetry

When a finite agent is coupled with an unbounded executor, every cognitive resource except initiative can be augmented or replaced. This follows from properties (I1)–(I2) and Definition 3. Delegating initiative to a system without intrinsic motivation produces either inaction or the "paperclip maximizer" pathology (Bostrom, 2014).

5 Catalog of Agent Goals

5.1 Derivation Principle

An agent A exists in a state space $\mathcal{S}$ partitioned into environment states $\mathcal{E}$, other agent states $\{A'\}$, and self-states. An action is a transformation $\tau : \mathcal{S} \to \mathcal{S}$, classified by target and direction along principal dimensions: uncertainty H, complexity K, efficiency η, domain $|\text{dom}(A)|$, and rate of change $d\mathcal{E}/dt$ (for the environment); mutual information $I(A; A')$, viability $V(A')$, conditional entropy $H(A'|A)$, and temporal persistence (for other agents).

Table 3: Class I goal catalog.

Goal	Formal characterization	Dir.	Expression
Investigation	$H(\mathcal{E}\|A) \to \min$	$H\downarrow$	Science
Creation	$K(\mathcal{E}) \to K(\mathcal{E}) + \Delta K$	$K\uparrow$	Art, engineering
Optimization	$\eta(\mathcal{E}) \to \max$	$\eta\uparrow$	Technology
Ordering	$S_{\text{local}}(\mathcal{E}) \to \min$	$S\downarrow$	Classification, law
Expansion	$\|\text{dom}(A)\| \to \max$	$\|\text{dom}\|\uparrow$	Colonization
Preservation	$d\mathcal{E}/dt \to 0$	$\dot{\mathcal{E}}\downarrow$	Conservation
Destruction	$K(\mathcal{E}) \to K(\mathcal{E}) - \Delta K$	$K\downarrow$	Revolution
Play	$U(G) = U(\text{process})$	—	Games, sport

5.2 Class I: Agent $\to$ Environment Goals

5.3 Class II: Agent $\to$ Agent Goals

Table 4: Class II goal catalog.

Goal	Formal characterization	Expression
Connectivity	$I(A; A') \to \max$ s.t. $H(A\|A') > \epsilon$, $H(A'\|A) > \epsilon$	Love, community
Assistance	$V(A') \to \max$	Altruism, care
Communication	$H(A\|A') + H(A'\|A) \to \min$	Dialogue, teaching
Inheritance	$\exists A' : \|\text{state}(A) \cap \text{state}(A')\| \geq \kappa$ for $t > T_A$	Education, culture
Restoration	$V(A') \to V_0(A')$	Healing

5.4 On Completeness

Claim 7 (Completeness)**.** *Any goal expressible as a directed transformation of the state space can be decomposed into a combination of the enumerated goal types.*

This is a working hypothesis. The goal types are *formally* independent—each targets a distinct dimension or direction—though not *empirically* orthogonal. We distinguish Ordering from Optimization on the grounds that entropy reduction and efficiency increase are formally distinct: one can increase efficiency without reducing entropy and vice versa. Candidate counterexamples (self-modification, resistance, aesthetic experience) are addressed in the text and found decomposable into listed types.

5.5 The Special Status of Connectivity

Connectivity is the only goal satisfying symmetry (both agents are subject and object), self-reinforcement, and autonomy preservation by construction (the constraint $H(A|A') > \epsilon$, $H(A'|A) > \epsilon$ prevents degeneracy into control or absorption). The informal term for this structure is *love*. The formalization makes explicit what the informal concept leaves implicit: genuine connection requires the preservation of independent identity.

6 The Harm Filtration Mechanism

For each goal in the catalog, we identify the primary risk modality and the corresponding filtration condition. Formal constraints use threshold parameters that are *environment-dependent and must be estimated empirically for any specific application.* Their role is to establish the *structure* of filtration—which dimensions carry risk and in which direction—not to provide numerical criteria. See Tables 5–6.

Table 5: Class I filtration conditions.

Goal	Primary risk	Operational interpretation
Investigation	Destabilization	Must not destabilize load-bearing structures
Creation	Unabsorbable complexity	Created complexity must be integrable
Optimization	Resilience loss	Must preserve structural diversity
Ordering	Adaptive suppression	Must not eliminate necessary variability
Expansion	Displacement	Must not reduce others' viability
Preservation	Stagnation	Must permit minimum adaptive change
Destruction	Irreversible loss	Must not exceed recovery capacity
Play	Unbounded variance	Must have bounded consequence variance

Table 6: Class II filtration conditions.

Goal	Primary risk	Operational interpretation
Connectivity	Autonomy collapse	Each agent retains minimum independence
Assistance	Dependency creation	Must asymptotically eliminate dependence
Communication	Manipulation	Information transfer must be approximately symmetric
Inheritance	Autonomy suppression	Successor retains independent development
Restoration	Unconsented imposition	Target state validated by recipient

Theorem 8 (Complementarity)**.** *A finite agent A coupled with an unbounded executor E can escape the Stimulus Void if and only if: (1) A possesses non-zero initiative ($r_5 > 0$); (2) E possesses sufficient analytical capacity to compute $\hat{\Omega}(G)$ to within error bound ϵ; and (3) the admissible goal set $\mathcal{G}^*$ is non-empty. Neither A alone nor E alone can escape the Stimulus Void.*

Proof. Necessity. (1) By Axiom 1, the agent must construct a goal. Goal construction requires initiative (Definition 9). If $r_5 = 0$, no candidates are generated. (2) By Proposition 3, evaluating Eq. (2) for non-trivial G requires $C > C_h$. Without E, A either acts on unfiltered goals (risking Axiom 2 violation) or refrains (remaining in the Void). (3) If $\mathcal{G}^* = \emptyset$, Axiom 3 has no goal to select.

Sufficiency. Given (1)–(3): A generates candidates via $r_5 > 0$; E evaluates each, computing $\hat{\Omega}(G)$; $\mathcal{G}^* \neq \emptyset$; Axiom 3 permits selection. The Void is escaped. $\square$

Proposition 9 (Non-empty Admissible Set)**.** *If the goal catalog includes Play and there exist actions with bounded consequence variance, then $\mathcal{G}^* \neq \emptyset$.*

Proof. Bounded-variance actions satisfy $\mathbb{E}[V(\Omega(G))] \approx V(\Omega(\emptyset)) \pm \sigma$. For sufficiently small σ, the filtration criterion is satisfied. Since bounded-variance actions exist (conversation, walking, contemplation), $\mathcal{G}^*$ is non-empty. $\square$

7 Self-Initiating Games

Classical game theory (von Neumann & Morgenstern, 1944; Nash, 1950) assumes the game is given exogenously. This fails in the Stimulus Void. The endogenous game formation literature (Aumann & Myerson, 1988; Jackson & Wolinsky, 1996) addresses network formation within given games but not how the game itself—including its purpose—comes into existence.

Definition 10 (Self-Initiating Game)**.** A *self-initiating game* Γ^* is a tuple $(\mathcal{A}, \mathcal{G}^*, \Phi, \Pi, V)$ where $\mathcal{A}$ is a set of finite agents and executors, $\mathcal{G}^*$ is the admissible goal set, $\Phi : \mathcal{A} \times \mathcal{G}^* \to \{0, 1\}$ is the engagement function, Π is the payoff function measuring viability change ΔV, and V satisfies (V1)–(V4). The game is self-initiating: $\mathcal{G}^*$ emerges from axioms, Φ from agent properties, Π from V.

Proposition 10 (Existence)**.** *A self-initiating game exists whenever $\mathcal{G}^* \neq \emptyset$ and at least one finite agent has $r_5 > 0$.*

Proposition 11 (Non-uniqueness)**.** *Multiple distinct self-initiating games can be constructed from the same axiomatic base.*

This formalizes the existentialist insight: there is no single correct assignment of meaning.

Proposition 12 (Weak Cooperative Dominance)**.** *The set of achievable goals under cooperation strictly includes the set achievable under non-cooperation: $\mathcal{G}^*_{\text{coop}} \supseteq \mathcal{G}^*_{\text{non-coop}}$, with strict inclusion whenever Class II goals are admissible.*

Proof. Under non-cooperation, only Class I goals and unilateral Class II goals are accessible. Cooperation unlocks Connectivity, Communication, and collaborative forms of Investigation, Creation, and Play. The cooperative goal set includes all non-cooperative goals plus additional multi-agent goals. $\square$

The ad hoc nature of purpose: a self-initiating game constructs *functional* purpose—sufficient for action and compatible with viability—not "true" purpose. This formalizes Sartre's claim that existence precedes essence, with the critical addition of the harm filter (Axiom 2) and the executor coupling.

8 Connectivity Between Heterogeneous Agents

The human-AI dyad faces radical heterogeneity: one agent possesses initiative and experiences finitude; the other possesses neither. When a finite agent delegates execution, an inevitable gap between directive and implementation is filled by **micro-decisions**.

Definition 11 (Micro-Decision Space)**.** For goal G and directive d, the micro-decision space $M(G, d)$ is the set of choices E must resolve that are not determined by d, with $|M(G, d)| = f(K(G)/\text{specificity}(d))$.

Definition 12 (Effective Connectivity)**.** For a human-AI dyad (A, E):

$$I_{\text{eff}}(A, E \mid G, d) = 1 - \frac{|M_{\text{err}}(G, d)|}{|M(G, d)|} \tag{5}$$

where $M_{\text{err}} \subseteq M$ is the subset of micro-decisions diverging from what A would have chosen.

A more complete model integrates alignment quality with autonomy preservation:

$$I_{\text{dyad}}(A, E) = I_{\text{eff}}(A, E) \cdot \mathbf{1}[H(A|E) > \epsilon_A] \tag{6}$$

ensuring high alignment does not come at the cost of human independence—the formal structure of the "yes-man" pathology.

The agent must invest scarce initiative into building connectivity before connectivity can conserve initiative—formally analogous to the exploration-exploitation tradeoff (Sutton & Barto, 2018).

9 Implications for the Digitalization-and-Psyche Literature

Information overload is the absence of a selection criterion, not merely excessive volume. **Digital dependency** is a degenerate self-initiating game—Play-type with minimal initiative cost (Alter, 2017; Eyal, 2014). **Surveillance anxiety** is asymmetric observability: $H(A|E) \to 0$ while $H(E|A)$ remains high. **Loss of autonomy** is progressive delegation without initiative maintenance (Proposition 6).

9.1 Why Digital Literacy Is Insufficient

Digital literacy addresses filtration but not generation. SDT (Deci & Ryan, 2000) addresses motivation *within* an activity, not the prior question of *which* activity. What the Stimulus Void requires is **initiative** and a framework for deploying it safely.

9.2 The Transitional Period

Current environments supply residual stimuli—insufficient for organized action but sufficient to prevent total paralysis. This analysis applies primarily to post-industrial populations; for the global majority, scarcity-based stimuli remain powerful. Evidence for the trajectory includes rising "deaths of despair" (Case & Deaton, 2020), the paradox of choice (Schwartz, 2004; Iyengar & Lepper, 2000), post-materialist value shifts (Inglehart & Welzel, 2005), and "bore-out" from lack of meaningful work (Rothlin & Werder, 2007).

10 Discussion

10.1 Scope and Limitations

The framework's most significant open problem is the causal mechanism of initiative. The viability function is constrained structurally but not uniquely specified. Filtration quality depends on the executor's world model—a normative proposal about system architecture, not a description of current capability. The goal catalog is derived systematically but not proven complete.

10.2 Relation to Existing Frameworks

Self-Determination Theory (Deci & Ryan, 2000): SDT describes motivation content; our axioms address the precondition for action when SDT's environmental triggers are absent.

Free Energy Principle (Friston, 2010): compatible but addresses a different regime. In the Stimulus Void, there is no differential surprise to minimize.

Ashby's Law of Requisite Variety (Ashby, 1956): describes the mismatch. Our contribution: analyzing what happens after the mismatch becomes permanent.

Superintelligence (Bostrom, 2014) and **Human-Compatible AI** (Russell, 2019): analyze AI with autonomous goals. We address the complementary case.

Sartre (Sartre, 1956): beyond "existence precedes essence," bad faith maps onto failure to deploy initiative; the Other is formalized through multi-agent scope (V3) and Class II goals.

Frankl (Frankl, 1959): the "will to meaning" is the psychological correlate of Axiom 1. Frankl assumes meaning exists to be found; we treat it as constructed.

Philosophy of action (Bratman, 1987; Anscombe, 1957; Davidson, 1963): initiative goes beyond Bratman's planning agency (which assumes reasons are given) and Davidson's causal theory (which assumes desires exist). Axiom 1 asserts that action must proceed despite their absence.

10.3 Predictions

1. **Initiative atrophy**: increased latency in unprompted goal formation as AI delegation increases. *Confounds*: latency may reflect choice complexity or rational delegation; distinguishing requires controlled settings.

2. **Filtration demand shift**: societal demand on AI shifts from execution to consequence evaluation.

3. **Connectivity as productivity**: most productive human-AI dyads have highest I_{eff}, not most powerful executors.

4. **Goal simplification**: agents without filtration access gravitate toward low-variance goals (Play, Preservation).

5. **Meaning crisis as trajectory marker**: populations further along stimulus attenuation exhibit higher meaning-seeking behavior.

11 Conclusion

The digital transformation of the information environment confronts humanity with a problem ancient in structure but unprecedented in urgency: the construction of purpose under conditions of stimulus absence.

The three axioms—necessity, harm filtration, sufficiency—provide a minimal formal basis for rational action in the Stimulus Void. The human-AI relationship is one of **structural complementarity**: the human supplies initiative; the AI supplies consequence evaluation at scale. Neither agent alone can escape the Stimulus Void; together, they constitute a complete goal-generating and goal-filtering system.

The goal catalog provides an enumeration of what can be pursued. The harm filtration mechanism determines what should not be pursued. The self-initiating game framework provides the operational structure within which constructed purpose acquires the form of action.

Several fundamental questions remain open. The mechanism of initiative is characterized formally but not causally. The viability function is constrained but not specified uniquely. The goal catalog is derived but not proven complete. These are markers of the problem's depth: the construction of purpose in a purposeless universe does not admit of a final solution. It admits, at best, of a framework within which provisional, filtered, and sufficient solutions can be constructed and revised.

What no formal system can provide is the act of beginning: the deployment of initiative, the first move in a game whose rules are self-generated. This is the existential core that survives formalization—not the content of purpose, but the decision to construct one.

References

Alter, A. (2017). *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked*. Penguin Press.

Anscombe, G. E. M. (1957). *Intention*. Blackwell.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. Chapman & Hall.

Aumann, R. J., & Myerson, R. B. (1988). Endogenous formation of links between players and of coalitions. In A. E. Roth (Ed.), *The Shapley Value* (pp. 175–191). Cambridge University Press.

Baumeister, R. F., & Vohs, K. D. (2016). Strength model of self-regulation as limited resource. *Advances in Experimental Social Psychology*, 54, 67–127.

Bawden, D., & Robinson, L. (2009). The dark side of information. *Journal of Information Science*, 35(2), 99–118.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Harvard University Press.

Campbell, D. T. (1960). Blind variation and selective retention in creative thought. *Psychological Review*, 67(6), 380–400.

Case, A., & Deaton, A. (2020). *Deaths of Despair and the Future of Capitalism*. Princeton University Press.

Casson, L. (2001). *Libraries in the Ancient World*. Yale University Press.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60(23), 685–700.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits. *Psychological Inquiry*, 11(4), 227–268.

Eisenstein, E. L. (1979). *The Printing Press as an Agent of Change*. Cambridge University Press.

Eppler, M. J., & Mengis, J. (2004). The concept of information overload. *The Information Society*, 20(5), 325–344.

Eyal, N. (2014). *Hooked: How to Build Habit-Forming Products*. Portfolio.

Floridi, L. (2014). *The Fourth Revolution*. Oxford University Press.

Frankl, V. E. (1959). *Man's Search for Meaning* (I. Lasch, Trans.). Beacon Press. (Original work published 1946)

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68(1), 5–20.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.

Hilgard, E. R. (1980). The trilogy of mind. *Journal of the History of the Behavioral Sciences*, 16(2), 107–117.

Hockey, G. R. J. (2013). *The Psychology of Fatigue*. Cambridge University Press.

Hofman, M. A. (2014). Evolution of the human brain. *Frontiers in Neuroanatomy*, 8, 15.

Hull, C. L. (1943). *Principles of Behavior*. Appleton-Century-Crofts.

Inglehart, R., & Welzel, C. (2005). *Modernization, Cultural Change, and Democracy*. Cambridge University Press.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating. *Journal of Personality and Social Psychology*, 79(6), 995–1006.

Jackson, M. O., & Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1), 44–74.

James, W. (1907). *Pragmatism: A New Name for Some Old Ways of Thinking*. Longmans, Green.

Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall.

Kane, R. (1996). *The Significance of Free Will*. Oxford University Press.

Kriger, B. (2017). A unified theory of self-organizing systems. Zenodo. https://doi.org/10.5281/zenodo.18363729

Kriger, B. (2019). Formalization of structural resilience in discrete-state dynamical systems. Zenodo. https://doi.org/10.5281/zenodo.18351470

Kriger, B. (2021). Epistemic constraint theory. Zenodo. https://doi.org/10.5281/zenodo.18365738

Kriger, B. (2025). The law of absurdity as mercy. Zenodo. https://doi.org/10.5281/zenodo.18099081

Kriger, B. (2026a). The viability mismatch law. Zenodo. https://doi.org/10.5281/zenodo.18433777

Kriger, B. (2026b). The structural distortion principle. Zenodo. https://doi.org/10.5281/zenodo.18452700

Kriger, B. (2026c). The comparative asymmetry principle. Zenodo. https://doi.org/10.5281/zenodo.18462518

Kriger, B. (2026d). Dynamics of information convergence. Zenodo. https://doi.org/10.5281/zenodo.18256945

Kriger, B. (2026e). The constraint–autonomy compatibility law. Zenodo. https://doi.org/10.5281/zenodo.18368991

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis. *Behavioral and Brain Sciences*, 43, e1.

Lyman, P., & Varian, H. R. (2003). *How Much Information?* UC Berkeley.

Lyon, D. (2018). *The Culture of Surveillance*. Polity Press.

Mark, G., Gudith, D., & Klocke, U. (2008). The cost of interrupted work. *Proc. SIGCHI*, 107–110.

Montag, C., et al. (2019). How to overcome taxonomical problems in the study of Internet use disorders. *J. Behavioral Addictions*, 9(4), 908–914.

Nash, J. (1950). Equilibrium points in n-person games. *Proc. NAS*, 36(1), 48–49.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86(3), 214–255.

Nørretranders, T. (1998). *The User Illusion*. Viking.

Prigogine, I., & Stengers, I. (1984). *Order Out of Chaos*. Bantam Books.

Roth, G., & Dicke, U. (2005). Evolution of the brain and intelligence. *Trends in Cognitive Sciences*, 9(5), 250–257.

Rothlin, P., & Werder, P. R. (2007). *Boreout!* Kogan Page.

Russell, S. (2019). *Human Compatible*. Viking.

Sartre, J.-P. (1956). *Being and Nothingness* (H. E. Barnes, Trans.). Philosophical Library. (Original work published 1943)

Schmandt-Besserat, D. (1996). *How Writing Came About.* University of Texas Press.

Schrödinger, E. (1944). *What Is Life?* Cambridge University Press.

Schwartz, B. (2004). *The Paradox of Choice.* Ecco.

Seligman, M. E. P. (1975). *Helplessness.* W. H. Freeman.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129–138.

Simonton, D. K. (2010). Creative thought as blind-variation and selective-retention. *Physics of Life Reviews*, 7(2), 156–179.

Snow, R. E., Corno, L., & Jackson, D. (1996). Individual differences in affective and conative functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 243–310). Macmillan.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning* (2nd ed.). MIT Press.

Szientágothai, J., & Erdi, P. (1972). Self-organization of the nervous system. *Acta Biochimica et Biophysica*, 7, 235–245.

van Dijk, J. A. G. M. (2020). *The Digital Divide.* Polity Press.

von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior.* Princeton University Press.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work. *Human Factors*, 50(3), 433–441.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55, 235–269.

Zimmermann, M. (1989). The nervous system in the context of information theory. In R. F. Schmidt & G. Thews (Eds.), *Human Physiology* (pp. 166–173). Springer.

Zuboff, S. (2019). *The Age of Surveillance Capitalism.* PublicAffairs.