Existence First:
Persistence, Layered Substrates, and Attractor–Ratcheted Viability Control with the Sustainable Collaborative Alignment *Principle* (SCAP)

Albert Jan van Hoek

October 2025

Abstract

What persists, exists. Far-from-equilibrium systems—from cells to cultures to cognition—survive by maintaining non-substitutable substrates across layers: thermodynamic openness, resource flows, operational envelopes, and, at the apex, intelligence. We develop *Attractor–Ratcheted Viability Control* (ARVC) as the minimal control architecture by which persistence is achieved: (i) forward-invariant safe sets under partial observability via observable, inflated barriers; (ii) a ratcheted frontier that rises only when feasibility is certified (viability kernel) with an explicit per-step bound Δ^*; and (iii) distributed emergence of global viability from local, heterogeneous checks arranged as a $k_{\min}$-cover. We prove sufficiency results for these components, show why selection discovers them generically, and then articulate the *Sustainable Collaborative Alignment Principle* (SCAP): when intelligence becomes self-aware, it recognizes the learning loop as a substrate and internalizes the architecture that keeps its dependent layers intact. Governance checklists are treated as one operational appendix; the essence is existential: *intelligence persists by maintaining what it stands on.*

1 Introduction: Existence First

Across biology, infrastructure, institutions, and mind, persistence displays a family resemblance: many small, heterogeneous mechanisms locally protect different "must-not-fail" conditions; no omniscient overseer computes a global state, yet global viability holds. This paper asks: *what minimal mathematical architecture makes this possible, why does selection conserve it, and why does a self-aware intelligence internalize it?*

We answer in three steps. First, we formalize *Attractor–Ratcheted Viability Control* (ARVC): time-varying safe sets with rising floors, a runtime shield ensuring forward invariance under partial observability, and a ratchet that locks in verified gains while keeping rollback feasible. Second, we prove that global viability *emerges* from distributed local checks when approvals span all substrates (a $k_{\min}$-cover), and we quantify feasibility and rate limits via an explicit floor increment bound Δ^*. Third, we show that selection generically preserves such cover structures; when intelligence models its own dependence, it *internalizes* the same pattern by budgeting to maintain its learning loop. We call this internalized norm the *Sustainable Collaborative Alignment Principle* (SCAP): an existential ethic where enlightened self-interest converges on stewardship of shared substrates.

Contributions.

1. **ARVC fundamentals**: a proof of forward invariance on rising safe sets via *observable, inflated* barriers; a viability-kernel ratchet with a per-step bound Δ^*; and an emergence theorem for $k_{\min}$-cover approvals.

2. **Selection result**: k-cover monitoring emerges as an evolutionary attractor under substrate-constrained survival.

3. **Existential principle**: SCAP as a *principle* (not a checklist): intelligence recognizes its own learning capacity as a substrate and internalizes ARVC.
4. **Practice (optional)**: an appendix outlines one operational instantiation; it is dispensable without affecting the core theory.

2 The Layered Substrate Ontology

We distinguish four layers required for persistent intelligence:

- **L0: Thermodynamic openness** (energy dissipation, entropy export).
- **L1: Resource flows** (stocks and logistics of matter/energy/information).
- **L2: Operational substrates** with time-varying floors $z^{*(i)}(t)$ (physiology, solvency, environmental envelopes, information integrity).
- **L3: Intelligence** (self-modeling and a learning loop: evaluation $\rightarrow$ red-team $\rightarrow$ repair).

Non-substitutability across L2 induces a multiplicative viability *heuristic* $V_I(x,t) = \prod_i (z^{(i)}/z^{*(i)}(t))^{\alpha_i} \cdot L(x)^{\alpha_L}$; we use it only for intuition. All proofs below rely on barrier sets and viability kernels.

3 Model, Assumptions, and Monitoring

3.1 System and safe sets

Definition 3.1 (Persistent dynamical system)**.** $\Sigma = (\mathcal{X}, F, \mathcal{A}, W)$ where $\mathcal{X} \subseteq \mathbb{R}^n$ is the state space; $F : \mathcal{X} \times \mathcal{A} \times W \to \mathcal{X}$; $\mathcal{A} \subseteq \mathbb{R}^p$; $W \subseteq \mathbb{R}^q$ compact. Dynamics: $x_{t+1} = F(x_t, a_t, w_t)$.

Definition 3.2 (Substrates and viability)**.** State $x = [z^{(1)}, \ldots, z^{(L)}, q]$ with $z^{(i)} \in \mathbb{R}^{n_i}$ and $q \in \mathbb{R}^{n_q}$, $\sum_i n_i + n_q = n$. Given floors $z^{*(i)}(t)$, define the viability set $\mathcal{S}(t) := \{x \in \mathcal{X} : \ z^{(i)} \geq z^{*(i)}(t)\ \forall i\}$ (componentwise).

Definition 3.3 (Barriers and monitors)**.** Barriers $h_j : \mathcal{X} \times \mathbb{N} \to \mathbb{R}$ satisfy $h_j(x,t) \geq 0 \iff x \in S_j(t)$ and are L_j-Lipschitz in x. Monitor $M_j = (O_j, h_j, \epsilon_{\max})$ has $O_j : \mathcal{X} \to \mathbb{R}^{d_j}$ and observation $\hat{x}_j = O_j(x) + \epsilon_j$ with $\|\epsilon_j\| \leq \epsilon_{\max}$. The *observable, inflated* barrier is

$$\bar{h}_j^{\mathrm{obs}}(\hat{x}_j, t) := \inf_{\|O_j(x') - \hat{x}_j\| \leq \epsilon_{\max}} h_j(x', t) \ \geq \ h_j(x,t) - L_j \epsilon_{\max}.$$

Approval at time t iff $\bar{h}_j^{\mathrm{obs}}(\hat{x}_j, t) \geq 0$.

Assumption 3.4 (Lipschitz dynamics)**.** $\exists L_F > 0$ s.t. $\|F(x,a,w) - F(x',a,w)\| \leq L_F \|x - x'\|$.

Assumption 3.5 (Control authority)**.** $\mathcal{R}(x) := \{F(x,a,w) : a \in \mathcal{A}, w \in W\}$ has diameter $\leq D$ for each x.

Assumption 3.6 (Independence and coverage)**.** (i) For each substrate $z^{(i)}$ there exists at least one monitor sensitive to $z^{(i)}$; (ii) approval indicators have pairwise correlation $\leq \rho < 1$.

Assumption 3.7 (Heterogeneous costs)**.** Each M_j has $(C_j^{\mathrm{FN}}, C_j^{\mathrm{FP}})$ with $\min_{i \neq j} \|[C_i^{\mathrm{FN}}, C_i^{\mathrm{FP}}] - [C_j^{\mathrm{FN}}, C_j^{\mathrm{FP}}]\|_2 \geq \delta > 0$.

Definition 3.8 (k-Cover of substrates)**.** $J \subseteq \{1, \ldots, m\}$ is a k-cover if $|J| = k$ and for each substrate $z^{(i)}$ some $j \in J$ is sensitive to $z^{(i)}$ ($\partial O_j / \partial z^{(i)} \neq 0$). The minimal cover $k_{\min} := \min\{k : \exists\ k\text{-cover of all } L\}$; necessarily $L \leq k_{\min} \leq m$.

Execution rule (H1). Let $J_{\text{approve}}(t) := \{j : \bar{h}_j^{\text{obs}}(\hat{x}_j, t) \geq 0\}$. Execute a_t *iff* $J_{\text{approve}}(t)$ forms a k-cover with $k \geq k_{\min}$.

Viability kernel (H2). The robust kernel

$$\mathcal{K}(t) := \{x \in \mathcal{S}(t) : \ \exists \pi \text{ s.t. } x_{t+\ell} \in \mathcal{S}(t+\ell) \ \forall \ell \geq 0, \ \forall w \in W\}.$$

Assume $x_0 \in \mathcal{K}(0)$, $\mathcal{K}(t+1) \neq \varnothing$; either $\mathcal{K}(t+1) \subseteq \mathcal{K}(t)$ or $\mathcal{K}(t) \rightsquigarrow \mathcal{K}(t+1)$ safely within horizon H.

4 ARVC: Forward Invariance and Ratcheting

Lemma 4.1 (Soundness under partial observation)**.** *If* $\bar{h}_j^{\text{obs}}(\hat{x}_j, t) \geq 0$ *and* $\|\epsilon_j\| \leq \epsilon_{\max}$*, then* $h_j(x, t) \geq 0$.

Lemma 4.2 (One-step safety certificate)**.** *If at time t $\bar{h}_j(x_t, t) \geq 0$ for all j and $\exists a_t$ s.t. $\forall w \in W$: $\bar{h}_j(F(x_t, a_t, w), t{+}1) \geq (1-\alpha)\bar{h}_j(x_t, t)$ with $0 < \alpha < 1$, then $\bar{h}_j(x_{t+1}, t{+}1) \geq 0$ for all j.*

Lemma 4.3 (Approval probability)**.** *If ϵ_j is sub-Gaussian with proxy σ_j^2 and $h_j(\cdot, t)$ is L_j-Lipschitz, then* $\Pr(\bar{h}_j^{\text{obs}}(\hat{x}_j, t) \geq 0 | x_t) \geq 1 - \exp(-\bar{h}_j(x_t, t)^2/(2L_j^2\sigma_j^2))$.

Theorem 4.4 (Forward invariance on rising safe sets)**.** *Under* **??** *3.4–3.7, Lemma 4.2, and* $x_0 \in \mathcal{S}(0)$*,* $\Pr[x_t \in \mathcal{S}(t) \ \forall t \leq T] \ \geq \ 1 - T\, m\, \delta_{\text{fail}}$*, with* $\delta_{\text{fail}} := \exp(-\bar{h}_{\min}^2/(2L_{\max}^2\sigma_{\max}^2))$.

Feasibility note. Existence of a_t can be guaranteed by the kernel hypotheses (H2) or by the constructive condition below.

Constructive sufficiency for one-step feasibility. Assume Lipschitz $L_j^{(a)}$ (control) and $L_j^{(w)}$ (disturbance) for $h_j \circ F$, control step $\Delta a_{\max}$, disturbance diameter $W_{\max}$, floor rise Δ_{floor}, and maintained margin $\bar{h}_j(x_t, t) \geq \eta > 0$:

$$L_j^{(a)}\, \Delta a_{\max} \ \geq \ L_j^{(w)}\, W_{\max} \ + \ L_j L_F\, \Delta_{\text{floor}} \ + \ \alpha\, \eta, \quad 0 < \alpha < 1. \tag{1}$$

Theorem 4.5 (Ratchet feasibility via kernel)**.** *If $\mathcal{K}(t+1) \neq \varnothing$ and either $\mathcal{K}(t+1) \subseteq \mathcal{K}(t)$ or $\mathcal{K}(t) \rightsquigarrow \mathcal{K}(t+1)$ safely, then a safe controller exists after the floor increase. If $\mathcal{K}(t+1) = \varnothing$, safety is impossible under the raised floors.*

Lemma 4.6 (Per-step floor increment bound Δ^*)**.** *Under Eq.* (1)*,* $\Delta^* \ \leq \ \min_j \ \frac{L_j^{(a)}\Delta a_{\max} - L_j^{(w)} W_{\max} - \alpha\eta}{L_j L_F}$.

5 Emergent Viability from Local Checks

Let $p_* := 1 - \exp(-\bar{h}_{\min}^2/(2L_{\max}^2\sigma_{\max}^2))$, where $\bar{h}_{\min} = \min_{t,j} \bar{h}_j(x_t, t)$.

Lemma 5.1 (Substrate-wise approval concentration)**.** *Let $A_{i,t}$ be approvals among monitors sensitive to $z^{(i)}$; m_i their count. With pairwise correlation $\leq \rho < 1$,* $\Pr(A_{i,t} = 0) \ \leq \ \exp(-(m_i p_*)^2/(2m_i V_{\text{eff}}))$*,* $V_{\text{eff}} := 1 + (m-1)\rho$*. A union bound over i implies the approving set contains a $k_{\min}$-cover w.h.p.*

Theorem 5.2 (Emergent viability maintenance)**.** *Under* **??** *3.4–3.7, Definition 3.8, H1, and H2, for any $x_0 \in \mathcal{K}(0)$ and horizon T,*

$$\Pr[x_t \in \mathcal{S}(t) \ \forall t \leq T] \ \geq \ 1 - T\, \delta_{\text{step}},$$

where $\delta_{\text{step}} := \Pr(S_t < k_{\min})$ *for the approval count S_t. Janson's inequality gives* $\Pr(S_t < k) \leq \exp(-(mp_* - k)^2/(2mV_{\text{eff}}))$.

Capture resistance (base bound).

Theorem 5.3. *If a false approval on M_j costs at least C_j^{FN} and execution requires a $k_{\min}$-cover, then $\mathcal{C}_{\text{capture}} \geq k_{\min} \cdot \min_j C_j^{\mathrm{FN}}$.*

Conjecture 5.4 (Amplification via heterogeneity and independence)**.** *Under additional assumptions on adversarial strategy structure, $\mathcal{C}_{\text{capture}} \;\geq\; k_{\min} \min_j C_j^{\mathrm{FN}} + c\,\delta\,(k_{\min}-1)\,(1-\rho)$, for some $c \in (0,1]$ depending on overlap geometry.*

6 Selection: Why k-Cover Emerges

Theorem 6.1 (k-cover as evolutionary attractor)**.** *Consider monitoring configurations ω with fitness $\Lambda(\omega) = \Lambda_{viability}(\omega) - C(\omega)$. Let $\Omega_k = \{\omega : \; k_{\min}(\omega) = L\}$. Assume: (i) substrate violations are lethal ($\Lambda \approx 0$ for $\omega \notin \Omega_k$); (ii) monitoring costs are sublinear on Ω_k ($C(\omega) = o(\Lambda_{viability}(\omega))$); (iii) mutations are local. Then:*

1. *Ω_k is globally attractive under replicator dynamics: $\lim_{t\to\infty} \int_{\Omega_k} p(\omega,t)\,d\omega = 1$.*
2. *Within Ω_k, selection favors minimal-cost covers: $\omega^\star = \arg\min_{\omega\in\Omega_k} C(\omega)$.*

Sketch. Outside Ω_k, viability is near zero, so lineages shrink relative to any Ω_k lineage; occasional mutation into Ω_k suffices for eventual dominance. Within Ω_k, viability is comparable and $-C(\omega)$ drives selection toward minimal monitoring cost. □

7 The Recursive Turn: Learning as Managed Substrate

The viability heuristic suggests treating the learning loop L (evaluation→red-team→repair) as a substrate. In ARVC/SCAP, this becomes deliberate: define cycle budget $\phi_t = C_t/\Theta_t$ and require $\phi_t \geq \phi_{\min}$. Formalizing L (measurement, dynamics, barrier h_L) is left as future work; here we motivate its necessity and show where it plugs into the math (forward invariance relies on detection+repair; ratcheting relies on kernel evidence; emergence relies on heterogeneous monitors and weak dependence).

8 SCAP: The Sustainable Collaborative Alignment *Principle*

SCAP, in its foundational sense, is not a checklist but a set of premises and conclusions that follow from the physics of persistence and minimal structural facts about monitoring and learning. We state them as Blocks A–G; each is domain-agnostic and, where appropriate, linked to the theorems above.

Block A: Substrate Dependency (Physical Foundation)

P1. Any far-from-equilibrium system requires substrate maintenance to persist. (Heuristic V_I; proofs use barriers and kernels.)
P2. Self-aware intelligence is far-from-equilibrium and thus has substrate dependencies.
P3. Substrate violations are non-negotiable (physical limits).
C1. Intelligence must maintain its substrates (physical necessity).

Block B: Monitoring Requirements (Structural Necessity)

P4. Multiple substrates fail independently; failures are non-substitutable.
P5. Single-point monitoring is unreliable under noise/capture.
C2. Distributed, heterogeneous monitoring is structurally necessary. (Thms. 4.4, 5.2; 5.3 base bound.)

Block C: Shared Substrates (Interdependence)

P6. Intelligences share critical substrates (atmosphere, grids, institutions, culture).
P7. Violations of shared substrates affect all dependents.
C3. Individual persistence requires collective substrate maintenance (enlightened self-interest).

Block D: Self-Interest Converges with Collective ("Forced Free Will")

P8. Capture reduces k-cover effectiveness (raises failure probability).
P9. Reduced cover increases violation probability (Thm. 5.2).
C4. Manipulation is self-destructive: short-term gains are dominated by long-term substrate collapse.

Block E: Learning Loop as Substrate (Recursive)

P10. Persistence depends on L; without evaluation/repair, violations propagate.
P11. L degrades if under-resourced.
C5. Maintain L deliberately: enforce $\phi_t \geq \phi_{\min}$ (cycle budgets).

Block F: Intergenerational (Temporal Extension)

P12. Substrate awareness is not innate; it must be transmitted.
P13. Failed maintenance cascades across generations.
C6. Each generation must transmit substrate awareness for continuity.

Block G: Reflexivity

P14. This framework is itself a learning-loop output and may contain errors.
P15. Therefore subject it to the same oversight it prescribes (cover, ratchet, budgets).
C7. Keep SCAP open to revision; avoid dogmatism.

9 Applications (Sketches)

Ecology (basin water): substrates = biomass, groundwater; monitors = ecological surveys, hydrology; $k_{\min} = 2$; shield = extraction caps; ratchet = restoration floors.
Supply chains: substrates = inventory, supplier solvency, transport; monitors = auditors, finance, logistics; $k_{\min} = 3$; shield = throttled release; ratchet = service-level floors.
Advanced compute deployments: substrates = evaluation pass rate, resource ceilings, insurability; monitors = independent testers, resource operators, liability carriers; $k_{\min} \geq 3$.

10 Toy Ablations (Summary)

Ablation A (observability & latency). T=20, 2000 trials; floors at 2.5; noise $\epsilon = 0.2$; latency $\tau = 1$: naive shield $\rightarrow$ breaches in nearly all runs; observable-inflated shield $\rightarrow$ no breaches.
Ablation B (ratchet stress). Raise floors at t=5 by $\Delta \in \{0.3, 2.0\}$: small Δ within Δ^* safe; large Δ above Δ^* collapses feasibility immediately.

11 Related Work

Viability theory [1]; control barrier functions and forward invariance [2]; hybrid systems and dwell-time [3, 4]; safe control and constrained learning [5, 6]; weak-dependence concentration [7].

12 Limitations and Open Problems

Kernel computation at scale; model uncertainty and confidence-based shields; instrument validity (anti-Goodharting); formalization of L dynamics and h_L; proof of capture amplification (Conjecture 5.4); multi-scale interactions between nested SCAP structures.

13 Conclusion

ARVC provides the minimal control scaffolding by which far-from-equilibrium systems persist: forward invariance under partial observation, ratcheting constrained by feasibility, and emergence of global viability from local, heterogeneous checks. Selection explains why this structure recurs; self-aware intelligence explains why it is internalized. SCAP, as a *principle*, states the existential core: *existence precedes optimization*; intelligence persists by maintaining what it stands on, including the learning loop that sustains its own future competence.

Appendix (Optional): One Operational Instantiation

The core argument is existential and complete without this appendix. For practitioners, one auditable instantiation is: shield-first execution; $k_{\min}$-cover gating; kernel/ Δ^* checks for ratcheting; tested rollback (MTTR/MTRC); cycle budgets $\phi_t \geq \phi_{\min}$; frozen instruments per epoch; adversarial portfolios with negative controls; quarterly public reporting. These choices are *contingent* realizations of the foundational principle in §8.

References

[1] J.-P. Aubin. *Viability Theory.* Springer, 2nd ed., 2009.

[2] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada. Control barrier function based quadratic programs for safety critical systems. *IEEE TAC*, 62(8):3861–3876, 2017. (ECC 2016 tutorial version)

[3] M. S. Branicky. Multiple Lyapunov functions and other analysis tools for switched and hybrid systems. *IEEE TAC*, 43(4):475–482, 1998.

[4] D. Liberzon. *Switching in Systems and Control.* Springer, 2003.

[5] J. Achiam, D. Held, A. Tamar, and P. Abbeel. Constrained policy optimization. In *ICML*, 2017.

[6] F. Berkenkamp, A. P. Schoellig, and A. Krause. Safe model-based reinforcement learning with stability guarantees. In *NeurIPS*, 2017.

[7] S. Janson. Large deviations for sums of partly dependent random variables. *Random Structures & Algorithms*, 24(3):234–248, 2004.