

Working paper — academic draft

Not peer-reviewed and not journal-published. Draft v5 (2026). Content may change. © José Luis D'Andrade · IdentIA / Self-Improving Society. **All rights reserved.** Brief citation with attribution is permitted. **Reproduction, translation, adaptation, or reuse**, in whole or in part (including commercial use), is **prohibited** without the **copyright holder's prior written permission**.

The Selection Ladder: An Ordering Constraint for Bounded Recursive Self-Improvement

Author: José Luis D'Andrade **Affiliation:** LegalDesign.DO **Project:** IdentIA — identia.do **Contact:** www.linkedin.com/in/josedandrade **Draft:** v5.0 (position) — position paper. Splits `paper-v4.md` (retired 2026-08-27; in git history). **Companion systems paper:** `paper-v5-systems.md` — *Persistent Agent Institutions as a Deployment Unit: Identity Invariants, Governance-as-Code, and a Measured Reliability Gate*. Cited below as **[PA]**. **Date:** August 2026 · **Revised:** 2026-09-28 (follows the retraction in [PA] §7.3) · **Spanish:** `paper-v5-position-es.md`

Main-text-only version, generated 2026-09-28 from `paper-v5-position.md`: no acknowledgements, AI statement, references or appendices. The `[n]` markers and the mentions of Appendix E point to the full file. Not committed.

Abstract

A system that can edit its own scaffolding needs a way to decide which edits survive. The dominant answer in the current literature is to let the system judge its own output — with a model, a rubric, or a benchmark — and keep what scores higher. We argue that this is the wrong default, and that the ordering of what a system must be able to *measure* is itself the safety property.

We present a six-rung **selection ladder** (F0–F5) in which recursive self-improvement is the *last* rung and is structurally unreachable until three prior properties hold: that fitness is **difficulty-adjusted** and measurably **reliable** (F2), that acting on it is **stable and reversible** (F3), and that **rare-but-valuable niches are protected** from being out-competed by an early winner (F4). Each rung advances on a **self-trigger** — a condition the system measures about its own data, not a date on a roadmap. The dependency is written as a conjunction the system evaluates about itself, and **Done**(*F*2) means the live measurement, not the existence of the code.

This is a position paper and the position has not been tested. F0–F2 are implemented and one of them has run live; that half is reported in the companion systems paper **[PA]**, which also **retracts** the live measurement that earlier versions of both papers offered as the gate blocking everything above it ([PA] §7.3): the convergence flag it rested on could not become `True` under the criterion then in use. F3 and F4 compute in shadow and have never

acted, **by decision rather than by a measured gate**. F5 has never run. The properties this paper calls preconditions are therefore **prerequisites we impose, not guarantees we have verified**, and §6 states what would falsify the position rather than leaving it unfalsifiable.

We also state, separately and as a normative claim rather than a result, the **metacentrality** thesis that shaped the governance scheme around the ladder: that every architecture of power tends to capture or to dissolve, and that the response is a center deliberately constrained by audit, rotation, transparency and self-limitation. We give it its own section, its own limits, and no borrowed credibility from the measurements in [PA].

Keywords: recursive self-improvement; quality-diversity; item response theory; selection pressure; AI governance; agent societies.

1. Introduction

The agent runtime already provides a **variation operator**: modern agents author and patch their own skills and consolidate their own memory. What the field has not settled is the **selection operator** — the procedure that decides which of those self-authored changes are actually better and should survive.

The prevailing pattern is to let the changed object certify itself: a model scores its own output, a benchmark number goes up, the edit is kept. This works until the signal is gameable, and every signal built from the same source as the thing being optimized is gameable. The failure mode is not dramatic — it is drift that reads as improvement.

Our position is that the fix is **ordering**, and that ordering is not a roadmap but a constraint. Before a system is allowed to edit itself, it must first be able to demonstrate — about its own data, by its own measurement — that the signal it would optimize is difficulty-adjusted, reliable, stable to act on, and not collapsing its own diversity. Each of those is a rung. None of them is novel as an estimator; every one is borrowed unchanged from an established literature. The claim is about what they are wired to **gate**.

We state this as a position because it is one. The ladder's lower rungs are built and one has run; the rungs that carry the argument have not. Section 6 is therefore not a formality: it is where we say what result would tell us this is wrong.

2. Related Work

2.1 Self-improving and self-designing agents

Voyager [4] built an ever-growing, executable **skill library** for a single Minecraft agent via an automatic curriculum and self-verification. **ADAS / Meta Agent Search** [5] has a meta-agent *program new agents in code* over a growing archive. The **Darwin Gödel Machine** [6] relaxes the original Gödel Machine's [26] proof requirement to *empirical* validation, iteratively rewriting its own code and keeping an archive of discovered

variants. **AlphaEvolve** [7] (the successor to **FunSearch** [8]) evolves programs by direct changes to their code, against automated evaluators. **PromptBreeder** [16], **EvoPrompt** [35] and PBT [15] evolve prompts and configurations.

SIS adopts these as the *variation* operator and the *evolutionary search* of its top rung (F5). Its distinct contribution is the **ordering discipline**: a *ladder* in which empirical fitness is made trustworthy *before* it is allowed to act, via difficulty adjustment (Item Response Theory, F2) and diversity preservation (MAP-Elites, F4), with an anti-Goodhart held-out set that is *computed but does not yet gate* (§7). The Darwin Gödel Machine demonstrates empirical fitness + archive but does not impose a difficulty-adjusted, self-triggered, governance-gated staging of *when* recursive editing is permitted. SIS makes that staging the central safety argument (§4), and — unlike a pure benchmark-driven evolver — measures the *reliability of its own fitness signal* (the IRT convergence and θ -stability self-triggers) before letting selection act.

Closest to SIS's framing is **Synergy** [31], which defines the *Agentic Citizen* of an "Open Agentic Web" by three requirements — web-native collaboration, agent identity and personhood, and lifelong evolution — and implements the third as *experience learning*: interaction trajectories are encoded with multi-dimensional dialogue-derived rewards and re-injected at inference time by value-aware retrieval, yielding large within-benchmark gains (+31–36% relative on SWE-bench Verified; +148% on OpenRCA) and immediate transfer to fresh instances. The convergence of questions with IdentIA is striking — Synergy's typed memory, profile and agenda parallel our SOUL/memory/lineage decomposition — but the selection discipline differs fundamentally: its dialogue-derived reward feeds retrieval directly, with no difficulty adjustment, no held-out anti-Goodhart anchor, and no reliability gate between measurement and action; governance is named as an open problem rather than provided as a mechanism. SIS occupies the complementary position: the *institution* as the unit that makes agentic citizens governable — selection acts only when measured reliable, and transmission of capability is gated by validation.

2.2 Selection, rating and quality-diversity

The estimators SIS climbs are classical. **Bradley–Terry** [12] models pairwise "better-than"; **TrueSkill** [13] gives Bayesian skill with uncertainty and draws; **Item Response Theory** [20] separates latent ability from item difficulty; **Thompson sampling** [22] turns posteriors into a routing policy; **NSGA-II** [23] handles multi-objective trade-offs; **MAP-Elites** [14] and **POET** [21] preserve and grow behavioral diversity; **PBT** [15] evolves configurations online. **Chatbot Arena** [24] popularized Bradley–Terry for LLM evaluation at scale. SIS's contribution is not these estimators but the *assignment* of each to a distinct fitness object (skill, profile, fact, repertoire, routing) and the *self-triggered progression* between them (§3), plus the engineering choice to implement them with **no third-party rating dependency** (Python standard library only), which keeps the selection layer auditable and free of supply-chain surface.

3. The ladder

IdentIA's principles A (Dynamic Striving), B (Absolute Continuity) and C (Emergent Pragmatism), the A → C bridge, and the governance levels L1–L8 named on the rungs are defined in [PA] §3 and §6.1. This paper uses the labels without restating them.

3.1 Variation vs. selection

The agent runtime already provides a **variation operator**: it authors and patches its own skills and consolidates memory in a self-improvement review. What is missing is a **measured selection operator**: a way to decide *which* of those changes are actually better and should survive, propagate (the A → C bridge), and anchor (principle B). Schematically:

RSI = $\underbrace{\text{Variation}}_{\text{exists (runtime)}} \circ \underbrace{\text{Selection}}_{\text{this ladder}}$ in a closed loop, editing the scaffold, gated by fitness.

Variation without selection is **drift**; selection without variation has nothing to choose from. The framing is deliberately Darwinian [28]: improvement is never *declared* by the object that changed; it is *selected* by measured comparison against alternatives in an environment. The ladder builds selection first, and only its top rung closes the recursive loop.

3.2 Five fitness objects

There is no single "fitness number." Five objects carry differently-shaped signal:

Object	Raw signal	Source table	Target estimator
Skill	usage_count , success_count	skill_publications	Beta posterior → Bradley–Terry/TrueSkill
Profile/agent	task pass/fail, heartbeats	Kanban task_runs , runtime_heartbeats	IRT (difficulty-adjusted) + graph reputation
Fact	supersession, citations	facts / society_facts	survival/supersession rate + peer-prediction
Repertoire	diversity of active skills/profiles	fitness_repertoire	Quality-Diversity (MAP-Elites)
Routing	who gets the next task	fleet_manager , Kanban dispatch	Thompson sampling over posteriors

3.3 The rungs and their self-triggers

Each rung is a milestone that declares an algorithm, the data it needs, a **self-trigger** (a condition the system measures about *itself* to advance), its dominant IdentIA principle, and its governance level. Self-trigger thresholds are environment-configurable (SOCIETY_FITNESS_*). Badges reflect the codebase as of June 2026:  implemented & tested;  implemented, in-progress or shadow;  design-only.



3.4 The measured rungs, in one paragraph

F0 places a Beta–Bernoulli posterior on each object's success rate and ranks by the **Wilson lower bound**, so that an object with one success in one attempt does not outrank one with ninety in a hundred — uncertainty is carried, not discarded. F1 converts outcomes into **relative** ability with Bradley–Terry [12] and TrueSkill [13], fed by a pairwise judge that evaluates each comparison in both orders to suppress position bias. F2 adds **Item Response Theory** [20]: ability θ is estimated jointly with item difficulty, so a profile cannot raise its rank by choosing easy work. All three are implemented and unit-tested, and F2 has run live; the mathematics, the code paths, the tests and the dated result are **[PA] §5.4–§5.6 and §7.2**. This paper does not restate them, because the argument below does not depend on their internals — only on the property F2 is built to certify, and on the fact that on the one occasion it ran, it reported that property as **absent**.

3.5 F3–F5 — acting, preserving diversity, and gated RSI

F3 (closed loop — implemented in shadow). Thompson sampling [22] over the Beta/TrueSkill/IRT posteriors selects the next assignee (explore/exploit without hand-tuning); **NSGA-II** [23] keeps routing on the Pareto front of {quality, cost, latency} via a fast non-dominated sort rather than maximizing quality alone; a Quality-Diversity coverage metric watches for monoculture. The engine (`evaluator/routing.py: thompson_select, non_dominated_sort, pareto_front, qd_coverage`) is unit-tested (`tests/test_routing_f3.py, 12/12`), and a **consumer is built into fleet-manager** (`routing_client, kanban_task_create`) but runs **shadow / disabled by default** (`SOCIETY_FITNESS_ROUTING_ENABLED=false`): it *recommends* but does not assign. Activation requires a live, converged F2 plus governance L3b/L5 and a real cost-signal source; the `active` mode is **fail-open** — it degrades to shadow if θ -stability or convergence is not met (`trigger_f3_ready`). Latency is already captured from `task_runs`; a real LLM-spend cost source is the missing input. Design and state: `docs/F3-F4-FLEET-WIRING.md`.

F4 (diversity — implemented in shadow). MAP-Elites [14] over a behavior descriptor (dominant domain $\times$ mean-difficulty band; cost tier deferred) keeps the elite of each cell (a per-cell ratchet), preventing *monoculture along the modelled descriptor dimensions*: rare-but-valuable niches are protected from being out-competed by the early winner. The archive (`migration 014_fitness_repertoire.sql, _recompute_repertoire, protected_elites, population_plateau`) is unit-tested (`tests/test_fitness_f4.py, 10/10`, incl. monotone coverage under selection pressure). A niche-protection consumer is built into profile retire/destroy paths (an elite requires `?force=true`; archival is *not* gated, per IdentIA-B) but is disabled by default and fail-open, awaiting F3-active wiring.

F5 (gated empirical RSI — design-only, frontier). Only here does the system edit its own scaffold (SOUL, skills) under the reliable fitness built below. **PBT** [15] exploit-and-explore over configurations — whose "exploit" step is the no-regression ratchet (principle B) — with **PromptBreeder** [16] or **EvoPrompt** [35] as the variation operator over SOUL/skills, and **POET** [21]-style generation of harder tasks when F4 detects a plateau. Variation is **always** gated by the F2 fitness. The RSI identity, moreover, must be a **dedicated persistent profile** (an `operator/self_improver` in a separate "Dev/Operator host" with the repository mounted behind `worktree/branch/test/build/smoke/governance` gates), not the conversational architect and not the runtime's default seed — because by principle B a self-improving identity needs its own SOUL, memory, sessions, lineage and observable fitness.

4. Why order is the claim

The three rungs of §3.5 are not a roadmap of features in the order we happened to build them. They are a dependency, and the dependency is the position this paper argues. Stated carefully: **the claim here is about necessary conditions, not demonstrated safety.**

F5 **cannot** be safe earlier, because only F2 makes fitness difficulty-adjusted (**designed to resist** gaming by easy-task selection) *and* measurably reliable (convergence + θ -stability); only F3 makes acting stable and reversible; and only F4 **is intended to** protect rare niches from being out-competed. None of those properties has been established empirically — F3 and F4 have never acted live and F5 has never run — so they are prerequisites we impose, not guarantees we have verified. Turning on recursive self-editing without F2–F4 would be "RSI without reliable selection = dangerous drift." The dependency is enforced:

$$\text{enable}(F5) \iff \text{Done}(F2) \wedge \text{Done}(F3) \wedge \text{Done}(F4).$$

Crucially, $\text{Done}(F2)$ is **not** "the IRT code exists"; it is the live self-trigger — stable θ *and* `converged=True`. Earlier versions of this paper added that the live run of June 2026 reported `converged=False`, and concluded that F3 activation, and therefore F5, was blocked by the system's own measurement rather than merely by policy. **[PA] §7.3 retracts that:** the flag could not become `True` under the criterion then in use, and the θ -stability half of the trigger satisfied itself on unchanged data. F3 and F5 are off, and as of today they are off **by decision**. The claim that the conjunction *has been observed doing the blocking* is withdrawn; the claim that the conjunction is the right thing to require is the argument of this paper, and is unaffected.

This stance has a classical antecedent. Gödel's second incompleteness theorem [27] shows that a sufficiently expressive formal system cannot establish its own consistency from within. We make no claim that the theorem applies literally to SIS; we adopt its engineering moral: the guarantees that authorize self-modification — convergence, θ -stability, the rotated held-out anchor, and ultimately the human gates of [PA] §6 — are deliberately placed *outside* the loop they certify. A self-improving system that grades its own improvement with the same signal it is optimizing has no ground to stand on; the ladder's external anchors are the ground.

5. Metacentrality: a normative position on where power should sit

The governance scheme expresses a design thesis: every architecture of power tends to **capture or dissolve**. The response is not to pick a pole but to design **conscious metacentrality** — a center with explicit audit, rotation, transparency and self-limitation. Operationally, five moves: recognize the tendency to capture; incorporate continuous internal verification; allow dynamic redistribution of power; maintain cohesion without absolutizing the center; integrate historical memory so cycles are not repeated unconsciously. These map onto the governance stack ([PA] §6.1), and **the mapping is uneven: two of the four moves run today and two do not**. Continuous internal verification is the evaluator and the **F0–F2** fitness layer (L4–L6), which are live; F3–F4 compute in shadow. Dynamic redistribution *would be* fitness-driven routing under budgets (L3b/L5) — but that routing is F3, disabled by default, and the L5 gate is enforced opt-in ([PA] §6.1), so this move is **stated, not operating**. Cohesion-without-absolutism **is** the no-unilateral-truth invariant plus quarantine (L4), enforced. Historical memory **is** archive-not-delete and fact supersession (B), enforced. As cognitive power outgrows human

political power, the defense is **metagovernance** (L7–L8) — verifiable limits and permanent audit that the center can apply to *itself* before becoming absolute — rather than more coercion. This is a normative position, offered for scrutiny, not an empirical result.

And it adds no constraint of its own. Every mechanism named above is enforced — or not — by the levels of [PA] §6.1 and the rungs of §3. Metacentrality *interprets* them; it does not add to them, and nothing in this section would fail if the section were deleted. What it does not specify is how the center audits *itself* beyond human approval, PR and tests: there is no rotation mechanism, no term limit, and no procedure by which L7–L8 could be turned against whoever holds them. That is the thesis's thinnest point. We name it here because the vocabulary of this section is more elevated than the governance table it interprets, and elevated vocabulary is exactly what makes an unenforced idea read as an enforced one.

6. What would falsify this position

A position that cannot be wrong is not a position. Four results would count against this one, in increasing order of how much they would cost us:

1. **Self-editing under an unreliable signal that does not degrade.** If a system were shown to edit its own scaffolding while its fitness signal was measurably unreliable — non-converged θ , unstable across re-estimation — and its downstream task performance did not degrade over a sustained run, the central claim weakens immediately. The ordering would then be a cost without a corresponding risk.
2. **Monoculture that does not occur.** F4's status as a precondition rests on the premise that unprotected selection collapses onto an early winner and loses rare-but-valuable niches. If a long-running society under F3-style routing *without* MAP-Elites preserved its repertoire anyway, F4 would be an optimization, not a gate, and the ladder would have one rung too many.
3. **A gate that never opens.** If `converged=True` proved unreachable at institutional data volumes — not because the design is wrong but because real traffic never produces the head-to-head density IRT needs — the ladder would be correct and useless. This is the failure we currently consider most likely, and it is the reason the gym exists ([PA] §7.1). One caution about how we would read it: until 2026-09-26 our own criterion could not report convergence for any data, so an unreachable flag is a claim about the data only once the instrument has been shown able to move ([PA] §7.3). It would not refute the ordering; it would refute the ladder as a practical instrument.
4. **The first live rung buying nothing.** The cheapest decisive experiment is the first item of [PA] §10's future work: reach `converged=True` on a client installation rather than the development environment where the convergence criterion was repaired, activate F3, and compare against round-robin routing on the same suite. If fitness-driven routing under a converged signal shows no measurable advantage over round-robin, the ladder's value proposition fails at the first rung where it could have been demonstrated.

We note which of these we can run and which we cannot. (4) is within reach and is the next piece of work. (1) and (2) require a longitudinal study of a society that has been allowed to act, which by construction cannot begin until (4) succeeds. (3) will only be answerable from production traffic over months.

7. Status: nothing in this paper has acted

The rungs this paper argues about have not run. Stated exactly, and traceable to the claims ledger in [PA] Appendix C:

- **F3** (Thompson sampling + NSGA-II routing) is implemented and unit-tested but runs **shadow, disabled by default**. It recommends; it does not assign. No path consults it as a precondition. Ledger E-14.
- **F4** (MAP-Elites archive, niche protection, plateau detection) is implemented and unit-tested and likewise **disabled by default and fail-open**, awaiting F3-active wiring. Ledger E-14.
- **F5** is **design-only and hard-gated**. It has never run.
- The reason none of them is active is a **decision, not a measurement**. Earlier versions said the opposite, citing the live F2 run of 2026-06-18 as a self-trigger that withheld activation; that reading is retracted ([PA] §7.3, ledger E-27), because the convergence flag could not become `True` under the criterion then in use. After the repair, a fit in a development environment reports `converged=True` (ledger E-28), and F3 stays off because we have not turned it on. Since `c811267` the flag is persisted and `fitness_route` in *active* mode does fall back to shadow when the latest fit did not converge, failing closed (ledger E-29) — built and tested, never exercised live.
- The anti-Goodhart held-out anchor, which the ladder's design treats as a guardrail, is **computed and served but blocks nothing** — an indicator, not a gate ([PA] §5.8, ledger E-15).

Of the 21 components tracked in [PA] §7.4, twelve do not currently constrain anything. The closed variation → selection → mutation loop needs items from two of those groups — F3 acting rather than shadowing, and F5 existing at all — so it cannot run today by construction, not by configuration.

8. Conclusion

We have argued that the ordering of measurable properties is the safety constraint in a self-improving agent society, and that recursive self-editing belongs at the end of that order rather than at the beginning of a roadmap. The ladder makes the dependency explicit and evaluable by the system about itself. We previously added that on the one occasion the gate ran it refused, and called that the only part of the argument observed rather than asserted; **[PA] §7.3 retracts it**, so no part of this argument has now been observed. The honest version is shorter: the ordering is written down, machine-checked as a rule, and untested in operation.

The rest is a position. F3 and F4 have never acted, F5 has never run, and the properties we call preconditions are prerequisites we impose. We have kept this argument out of the systems paper for that reason, stated in §6 what would tell us it is wrong, and named in §5 the thinnest part of the governance thesis that accompanies it — that it does not yet specify how the center audits itself. The useful next result is not a larger claim; it is the gym matrix redesigned until θ is identifiable, and then the first honest comparison between fitness-driven routing and round-robin.