Beyond Artificial Intelligence: The Mathematical Dynamics of General Intelligence

Recursive State-Space Dynamics, Self-Modeling, and Adaptive Goal Formation

From Sector-Gate Dynamics to a Formal Mechanics of AGI

Peter De Ceuster

August 27, 2026

Addressed to: Prof. Marcus Hutter (`marcus.hutter@anu.edu.au`)

Abstract

Thanks to the hard work of the community the mathematics of AI and AGI have now been established. It is now time to focus on design, this paper will provide an angle to design AGI. While modular sector-gate dynamics and quantum state representations establish a candidate mathematical substrate for multi-domain routing [1, 2], treating quantum hardware directly as cognition remains a working hypothesis without an overarching, substrate-independent definition of general intelligence [1]. In this paper, we shift the core theoretical question from *"how quantum hardware might implement AGI"* to *"what mathematical object constitutes an AGI."*

We formulate **General Intelligence Dynamics (GID)**, a formal mechanics of intelligence defined by core dynamical mechanisms over an augmented recursive state tuple $X_t = (B_t, M_t, G_t, S_t, W_t)$, spanning environmental belief states, internal memory structures, structured goal topologies, self-referential models, and working computational scratch spaces. Under GID, cognition is formalized not as passive information routing or static reward optimization, but as an active, recursive dynamical loop governed by coupled equations of motion:

$$\begin{aligned}
B_{t+1} &= \mathcal{B}(B_t, a_t, O_{t+1}, W_t),\\
M_{t+1} &= \mathcal{M}(M_t, O_{t+1}, X_t),\\
S_{t+1} &= \mathcal{S}(S_t, X_t, M_t, O_{t+1}),\\
G_{t+1} &= \Gamma(G_t, S_t, M_t, X_t),\\
W_{t+1} &= \mathcal{W}(W_t, X_t, a_t),\\
a_t &= \Pi(X_t, \theta_t),\\
\theta_{t+1} &= \mathcal{L}(\theta_t, X_t, O_{t+1}, G_t).
\end{aligned}$$

We mathematically formalize: (1) closed-loop prediction-action-observation recursion separating external environmental states $e_t \in \mathcal{E}$ from internal belief distributions B_t, (2) state-dependent internal transition operators $G_{ij}(X_t)$ that replace static circuits with belief-directed modular routing, building upon cross-sector interference models [1], (3) parametric learning dynamics yielding non-stationary transition kernels $P(X_{t+1}|X_t, O_{t+1}; \theta_t)$, (4) quantitative self-modeling mappings $S_{t+1} = \mathcal{S}(S_t, X_t, M_t, O_{t+1})$ calibrated by self-performance divergence $\mathcal{E}_S(t) = D_{\mathrm{KL}}(p_t \,\|\, \hat{p}_t)$, (5) explicit structured goal-generation operators $\Gamma : \mathcal{G} \times \mathcal{S} \times \mathcal{M} \times \mathcal{X} \to \mathcal{G}$ over tuples (g_i, c_i, w_i) and prove Pareto-optimal goal reconciliation on capacity-constrained domains, (6) formal cross-domain transfer operators $\mathcal{T}(\tau_i \to \tau_j)$ and a counterfactual generality functional $\mathcal{G} = \mathbb{E}_{(\tau_i,\tau_j)\sim\mathcal{D}^2_{\neq}}[P_j(\theta^{(i)}) - P_j(\theta^{(0)})]$ over ordered distinct task pairs, and (7) latent abstraction operators $\mathcal{A} : \mathcal{D}^k \to \mathcal{Z}$ capturing shared structural invariances across task distributions.

Finally, we establish a conceptual architectural hierarchy mapping these abstract mechanics through computational modules down to candidate physical substrates (such as topological quantum lattices [3] and verified execution frameworks [5]). This decouples the formal mathematics of general intelligence from its physical substrate while providing the mathematical foundation required for subsequent hardware compilation.

1 To AGI or not to AGI, that is the question.

The formal synthesis of Artificial General Intelligence (AGI) has historically vacillated between two disconnected extremes: heuristic software architectures lacking plausible mathematical foundations, and physical hardware analogies—such as quantum neural models—that suggest structural similarities to cognitive dynamics without formalizing the underlying mathematical object of intelligence [1, 2].

In prior foundational work, we introduced *Quantum Eigenstate Dynamics (QED)* [2] to formalize multi-variable cognitive superposition and collapse, allowing classical bits to be subsumed by qubit states enabling simultaneous resolution of binary logic via eigenvalue collapse [2]. This was followed by *Quantum Sector-Gate Dynamics (QSGD)* [1], which decomposed a global Hilbert space into modular cognitive sectors coupled by cross-sector unitary operators. While QSGD established rigorous algebraic tools—such as phase-aligned commutation invariants and localized Baker-Campbell-Hausdorff (BCH) corrections [1]—it explicitly acknowledged a fundamental caveat: mapping quantum transitions directly to cognitive processes remains a working hypothesis rather than an established physical mechanism until grounded in an overarching formal mechanics [1].

The purpose of this paper is to address this missing abstraction layer. We shift the core theoretical inquiry away from the hardware-first question:

"How can quantum computers implement AGI?"

and elevate it to the substrate-independent mathematical question:

What mathematical object is an Artificial General Intelligence?

Hence, in this work we contribute to the design of AGI. By defining intelligence as an autonomous dynamical system operating over recursive, self-referential state spaces, we propose a formal mechanics termed **General Intelligence Dynamics (GID)**. Under this framework, physical substrates (whether classical digital computers, neuromorphic networks, or topological quantum processors [3]) serve as execution layers for an abstract mathematical mechanics.

1.1 The Four-Paper Research Program

This treatise represents the critical theoretical core (Paper III) within an integrated four-stage research progression:

1. **Paper I — Quantum Eigenstate Dynamics for AGI Synthesis** [2]: Establishes continuous superposition, non-binary uncertainty encoding, and projection-driven resolution.

2. **Paper II — Quantum Sector-Gate Dynamics for AGI** [1]: Establishes modular sector factorizations, inter-sector couplings, and algebraic operator commutation principles.

3. **Paper III — A Mathematical Theory of General Intelligence (Present Work)**: Formalizes the mathematical mechanics of general intelligence independently of hardware, establishing the state space, self-modeling calibration, dynamic goal revision, cross-domain transfer, latent abstraction, and the GID state equations.

4. **Paper IV — Interferometric Sector-Gate Quantum Dynamics for AGI** [3, 5]: Provides the hardware-compilation mapping from abstract GID operators onto topological Majorana parity-interferometer lattices and verified runtime execution layers.

—

2 The Formal State Space and Causal Dynamical Loop

Classical reinforcement learning and algorithmic decision theory typically formalize an agent via discrete Markov Decision Processes (MDPs) characterized by environmental state $s \in \mathcal{S}$ and scalar reward optimization $\max \mathbb{E}[R(s,a)]$. General intelligence, however, operates under partial observability, non-stationary task distributions, internal resource constraints, and continuous self-reconfiguration.

2.1 Separation of Environmental State and Belief Manifold

Let $e_t \in \mathcal{E}$ denote the true, latent state of the external environment at discrete time $t \in \mathbb{N}$, governed by transition law $\mathcal{P}_{\text{env}}(e_{t+1} \mid e_t, a_t)$. The agent receives sensory observations $O_t \in \mathcal{O}$ generated via an observation emission channel $O_t \sim \Omega(\cdot \mid e_t)$.

Due to partial observability, the agent does not possess direct access to e_t. Instead, it maintains a structured belief state $B_t \in \mathcal{B}$, representing a filtered posterior distribution over environmental configurations:

$$B_t = \mathbb{P}\left(e_t \mid O_{\le t}, a_{<t}\right) \in \Delta(\mathcal{E}) \tag{1}$$

where $\Delta(\mathcal{E})$ denotes the probability simplex over environmental configurations.

2.2 State-Space Decomposition without Circularity

To eliminate mathematical circularities between active working states and committed actions, we delineate the internal cognitive state from the external action operator.

Definition 1 (General Intelligence State Space)**.** *The state of an AGI system at discrete time $t \in \mathbb{N}$ is formalized as an augmented 5-tuple X_t living in a composite product state space $\mathcal{X} = \mathcal{B} \times \mathcal{M} \times \mathcal{G} \times \mathcal{S} \times \mathcal{W}$:*

$$X_t = (B_t, M_t, G_t, S_t, W_t) \tag{2}$$

where:

- $B_t \in \mathcal{B} \subseteq \Delta(\mathcal{E})$ *represents the belief state and predictive estimate of the external environment given sensory history.*
- $M_t \in \mathcal{M}$ *represents the structured internal memory manifold (declarative, episodic, procedural, and associative representations).*
- $G_t \in \mathcal{G}$ *denotes the active structured goal topology and constraint network.*
- $S_t \in \mathcal{S}$ *is the explicit self-model, representing the system's recursive estimation of its own internal capacities, uncertainty bounds, and behavioral policies.*
- $W_t \in \mathcal{W}$ *denotes the working computational scratch space (registers, intermediate inference traces, and deliberative superposed amplitudes) [2].*

$$\boxed{\textbf{State } X_t \xrightarrow{\Pi} \textbf{Action } a_t \xrightarrow{\mathcal{E}} \textbf{Observation } O_{t+1} \xrightarrow{\mathcal{F}} \textbf{State Update } X_{t+1}}$$

Figure 1: The closed-loop causal dynamical trajectory of General Intelligence Dynamics.

The decision policy is a separate operator mapping the state space into an executable action $a_t \in \mathcal{A}_{\text{ext}}$:

$$a_t = \Pi(X_t, \theta_t) \tag{3}$$

yielding the clean causal sequence shown in Figure 1:

$$X_t \overset{\Pi}{\to} a_t \overset{\mathcal{E}}{\to} O_{t+1} \overset{\mathcal{F}}{\to} X_{t+1} \tag{4}$$

—

3 Adaptive Internal Transition Dynamics

In static computational models, internal state transitions are governed by rigid circuits or fixed weight tensors. In contrast, Sector-Gate dynamics [1] decomposes the active computational manifold into N interacting subspaces:

$$\mathcal{H}_{\text{AGI}} = \bigotimes_{k=1}^{N} \mathcal{H}_k \tag{5}$$

where $\mathcal{H}_k$ corresponds to functional sectors (e.g., perception $\mathcal{H}_1$, reasoning $\mathcal{H}_2$, memory $\mathcal{H}_3$, self-monitoring $\mathcal{H}_4$, and goal execution $\mathcal{H}_5$) [1, 2].

In static quantum circuits, a cross-sector operator $G_{ij} : \mathcal{H}_i \to \mathcal{H}_j$ applies a fixed unitary rotation [1]. In GID, internal transitions are generalized into *state-dependent nonlinear operator fields*.

Definition 2 (State-Dependent Internal Sector Operator). *Let G_{ij} be an operator inducing an information-theoretic transition from sector i to sector j. In GID, G_{ij} is parameterized dynamically by the full cognitive state tuple X_t:*

$$G_{ij} = G_{ij}(X_t) \equiv G_{ij}(B_t, M_t, G_t, S_t, W_t) \tag{6}$$

This parameterization ensures that the internal routing of representations across cognitive sectors is belief-directed. The system does not execute a predetermined sequence of transformations; rather, its active hypothesis S_t and goal configuration G_t modulate the operator generators:

$$G_{ij}(X_t) = \exp\left(-i \sum_m \theta_m(X_t) \hat{P}_i^{(m)} \otimes \hat{Q}_j^{(m)}\right) \tag{7}$$

where $\hat{P}_i$ and $\hat{Q}_j$ are sectoral projection or coupling operators [1, 5], and $\theta_m(X_t)$ is a state-dependent control functional. For conceptual illustration, an engineered sequence mapping perception through reasoning to memory functions analogously to a CNOT-style cross-sector gate, allowing a target sector to flip its state conditionally based on a control sector, thus physically encoding causal cognitive flow [1].

Proposition 1 (Phase Alignment and Dynamic Commutation). *Let two composite transitions act on overlapping sectors: $U = G_{jk}G_{ij}$ and $V = G_{ij}G_{jk}$. The ordering invariance $[U, V] = 0$ established in QSGD [1] holds dynamically if and only if the control functional satisfies:*

$$\theta_m(X_t) \equiv \theta_n(X_t) \pmod{2\pi} \tag{8}$$

across all basis manifolds permuted by downstream sector gates [1].

Proof. Following the matrix derivations in [1], the commutator between a diagonal controlled-phase sector gate and a permutation gate (such as CNOT) is supported exclusively on the swapped computational subspace:

$$\mathcal{C} = \left(e^{i\theta_m(X_t)} - e^{i\theta_n(X_t)}\right) (|110\rangle\langle 111| - |111\rangle\langle 110|) \tag{9}$$

When the control functional aligns the phases ($\theta_m(X_t) = \theta_n(X_t)$), the commutator vanishes identically ($\mathcal{C} = 0$), guaranteeing that belief-directed routing executes deterministically regardless of execution ordering [1]. □

—

4 Parametric Learning and Non-Stationary Dynamics

General intelligence requires that the system's transition laws adapt under experience. A purely unitary or static deterministic circuit cannot learn; learning requires non-unitary, dissipative adaptation of the parameters that govern internal routing.

Definition 3 (Parametric Learning Law)**.** *Let $\theta \in \Theta$ denote the coordinate parameters defining the transition operators and state evolution maps. The learning dynamics over discrete intervals is defined by:*

$$\theta_{t+1} = \mathcal{L}(\theta_t, X_t, O_{t+1}, G_t) = \theta_t + \eta \nabla_\theta \mathcal{F}_{adapt}(X_t, O_{t+1}, G_t) \tag{10}$$

where $\eta > 0$ is an adaptive learning rate tensor, and $\mathcal{F}_{adapt}$ is a variational functional evaluating predictive accuracy, energetic cost, and goal advancement.

Under continuous parameter updates, the internal conditional transition kernel becomes explicitly non-stationary:

$$P(X_{t+1} \mid X_t, O_{t+1}; \theta_t) \neq P(X_{t+1} \mid X_t, O_{t+1}; \theta_{t-\tau}) \tag{11}$$

This non-stationarity is regularized via information-preserving boundaries. When parameterized via unitary exploration manifolds U_θ, information gradients are preserved without premature lossy bottlenecks [4]:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{F}_\beta(U_\theta, C, M) \tag{12}$$

balancing model description length, interventional likelihood, and task performance [4]. This objective is structured so that any self-rewrite or parametric update must pass algorithmic invariant checks and maintain a tight boundary before its adoption is committed [4].

—

5 Quantitative Self-Modeling and Calibration

GID hypothesizes that general intelligence is characterized by the coupled presence of adaptive world modeling, transferable representations, self-model calibration, structured goal dynamics, and closed-loop action. While narrow systems construct world models $\hat{e} = f(e, a)$, an AGI must construct an explicit, quantitatively calibrated model of itself operating inside the world [4].

5.1 Non-Circularity in Self-State Estimation

To avoid circular definitions where a self-model is defined in terms of its instantaneous self, we delineate the internal prediction of the state from its recursive update.

Definition 4 (Self-Model Estimation and Update)**.** *Let $\hat{X}_{t+1} \in \mathcal{X}$ denote the system's prior estimate of its next state configuration. The self-model evolves via a non-circular update map $\mathcal{S}$:*

$$\hat{X}_{t+1} = \mathcal{S}_{est}(S_t, X_t, M_t) \tag{13}$$
$$S_{t+1} = \mathcal{S}(S_t, X_t, M_t, O_{t+1}) \tag{14}$$

incorporating external verification from observation O_{t+1}.

The presence of S_t enables counterfactual internal deliberation:

$$\text{Instead of querying only:} \quad \mathbb{P}(B_{t+1} \mid B_t, a_t) \tag{15}$$

$$\text{The system queries:} \quad \mathbb{P}(X_{t+1}, S_{t+1} \mid X_t, \mathrm{do}(W_t = \omega)) \tag{16}$$

$$\boxed{X_t \longrightarrow S_t \longrightarrow G_{t+1} \longrightarrow a_t \longrightarrow X_{t+1}}$$

Figure 2: The Meta-Cognitive Feedback Loop under Self-Modeling.

Definition 5 (Self-Model Calibration Error). *Let $\hat{p}_t(a) = \mathbb{P}(success \mid S_t, a)$ denote the system's subjective prediction of its operational success given self-model S_t, and let $p_t(a)$ denote its true empirical performance distribution. The self-model calibration error $\mathcal{E}_S(t)$ is defined as the Kullback-Leibler divergence:*

$$\boxed{\mathcal{E}_S(t) = D_{KL}\left(p_t(a) \,\|\, \hat{p}_t(a)\right) = \sum_{a \in \mathcal{A}_{ext}} p_t(a) \ln\left(\frac{p_t(a)}{\hat{p}_t(a)}\right)} \tag{17}$$

Proposition 2 (Conditional Calibration Convergence). *Assume that $\mathcal{E}_S(S)$ is L-smooth with respect to the parameterization of S_t, that $p_t(a)$ is quasi-stationary over calibration intervals, and that the stochastic gradient estimator $g_t = \widehat{\nabla}_S \mathcal{E}_S(t)$ has bounded variance $\mathbb{E}[\|g_t - \nabla \mathcal{E}_S(t)\|^2] \le \sigma^2$. If the update step $\Delta S_t = -\eta_s g_t$ satisfies $0 < \eta_s < \frac{2}{L}$, then the expected calibration error satisfies:*

$$\mathbb{E}\left[\mathcal{E}_S(t+1) \mid S_t\right] \le \mathcal{E}_S(t) - \eta_s\left(1 - \frac{L\eta_s}{2}\right)\|\nabla \mathcal{E}_S(t)\|^2 + \frac{L\eta_s^2\sigma^2}{2} \tag{18}$$

yielding an expected decrease outside a variance neighborhood of radius $\mathcal{O}(\eta_s\sigma^2)$ determined by the gradient and noise balance.

Proof. By L-smoothness of $\mathcal{E}_S$, the standard descent inequality yields:

$$\mathcal{E}_S(S_{t+1}) \le \mathcal{E}_S(S_t) + \langle \nabla \mathcal{E}_S(S_t), S_{t+1} - S_t \rangle + \frac{L}{2}\|S_{t+1} - S_t\|^2 \tag{19}$$

Substituting $S_{t+1} - S_t = -\eta_s g_t$ and taking the conditional expectation with $\mathbb{E}[g_t] = \nabla \mathcal{E}_S(S_t)$ and $\mathbb{E}[\|g_t\|^2] = \|\nabla \mathcal{E}_S(S_t)\|^2 + \sigma^2$:

$$\mathbb{E}[\mathcal{E}_S(S_{t+1}) \mid S_t] \le \mathcal{E}_S(S_t) - \eta_s\|\nabla \mathcal{E}_S(S_t)\|^2 + \frac{L\eta_s^2}{2}\left(\|\nabla \mathcal{E}_S(S_t)\|^2 + \sigma^2\right) \tag{20}$$

$$= \mathcal{E}_S(S_t) - \eta_s\left(1 - \frac{L\eta_s}{2}\right)\|\nabla \mathcal{E}_S(S_t)\|^2 + \frac{L\eta_s^2\sigma^2}{2} \tag{21}$$

For any $\eta_s \in (0, 2/L)$, the coefficient $1 - \frac{L\eta_s}{2} > 0$, guaranteeing strict expected reduction whenever $\|\nabla \mathcal{E}_S(S_t)\|^2 > \frac{L\eta_s\sigma^2}{2 - L\eta_s}$. $\square$

—

6 Explicit Goal Representation and Pareto Goal Dynamics

Standard machine learning assumes a fixed external objective functional $R(s, a)$. However, general intelligence in open-world environments requires dynamic goal autonomy: constructing subgoals, evaluating conflicts between competing objectives, and revising goal structures when encountering novel domains.

Definition 6 (Structured Goal Topology). *An active goal state $G_t \in \mathcal{G}$ is a discrete-continuous collection of N_t structured goal specifications:*

$$G_t = \left\{(g_i, c_i, w_i)\right\}_{i=1}^{N_t} \tag{22}$$

where:

- $g_i : \mathcal{B} \times \mathcal{W} \to \mathbb{R}$ *is the objective performance functional.*
- $c_i : \mathcal{X} \to \{0, 1\}$ *represents boundary constraints and safety invariants.*
- $w_i \in [0, 1]$ *represents the dynamic priority weight, normalized such that* $\sum_{i=1}^{N_t} w_i = 1$.

Definition 7 (Goal Evolution Operator)**.** *The goal state evolves via a dynamic operator* $\Gamma : \mathcal{G} \times \mathcal{S} \times \mathcal{M} \times \mathcal{X} \to \mathcal{G}$*:*

$$G_{t+1} = \Gamma(G_t, S_t, M_t, X_t) \tag{23}$$

encompassing six formal operations:

1. ***Goal Creation****: Spawning new goal tuples* $(g_{new}, c_{new}, w_{new})$ *when novel environmental regularities are identified.*
2. ***Goal Deletion****: Pruning satisfied or degenerate objectives.*
3. ***Goal Decomposition****: Partitioning complex objectives into hierarchical subgoals* $g_i \mapsto \{g_{i,1}, g_{i,2}, \ldots\}$.
4. ***Conflict Evaluation****: Detecting contradictory gradient flows between competing goals.*
5. ***Goal Revision****: Adjusting constraints* c_i *and priorities* w_i *based on self-model capacity* S_t.
6. ***Subgoal Generation****: Generating short-horizon operational targets supporting long-horizon convergence.*

The operator Γ executes an intrinsic cognitive cycle:

$$\boxed{\textbf{Generate} \longrightarrow \textbf{Evaluate} \longrightarrow \textbf{Reconcile Conflicts} \longrightarrow \textbf{Revise} \longrightarrow \textbf{Pursue}}$$

Definition 8 (Capacity-Constrained Admissible Goal Set and Pareto Frontier)**.** *Let* $\mathcal{G}_{feasible} \subset \mathcal{G}$ *denote the compact convex set of goal parameters satisfying hard boundary invariants* $c_i(X) = 1$. *For a given self-model* S_t *and divergence bound* $\delta > 0$*, define the capacity-constrained admissible goal set:*

$$\mathcal{G}_{adm}(S_t, \delta) = \left\{ g \in \mathcal{G}_{feasible} \;\middle|\; D_{KL}\left(S_t(g) \,\middle\|\, S_t(G_t)\right) \leq \delta \right\} \tag{24}$$

The Pareto frontier $\mathcal{P}(\mathcal{G}_{adm})$ *over* K *competing continuous utility objectives* $\boldsymbol{u}(g) = (u_1(g), \ldots, u_K(g))$ *is defined as:*

$$\mathcal{P}\left(\mathcal{G}_{adm}\right) = \left\{ g \in \mathcal{G}_{adm} \;\middle|\; \nexists g' \in \mathcal{G}_{adm} \text{ s.t. } u_k(g') \geq u_k(g)\ \forall k \text{ and } \exists j, u_j(g') > u_j(g) \right\} \tag{25}$$

Lemma 1 (Pareto-Optimal Goal Synthesis under Self-Model Constraints)**.** *Assume that each utility objective* $u_k : \mathcal{G}_{feasible} \to \mathbb{R}$ *is concave and continuously differentiable, and that* $\mathcal{G}_{adm}(S_t, \delta)$ *is convex with a non-empty interior. Then, for any strictly positive priority weight vector* $\boldsymbol{w} \in \mathbb{R}^K_{>0}$ *with* $\sum_{k=1}^{K} w_k = 1$*, the optimal solution* g^* *to the scalarized program:*

$$g^* = \arg\max_{g \in \mathcal{G}_{adm}(S_t, \delta)} \sum_{k=1}^{K} w_k u_k(g) \tag{26}$$

is strictly Pareto-optimal on the admissible set, i.e., $g^* \in \mathcal{P}(\mathcal{G}_{adm}(S_t, \delta))$*. Furthermore, if the trust-region self-model penalty is scalarized with parameter* $\lambda > 0$*:*

$$g^*_\lambda = \arg\max_{g \in \mathcal{G}_{feasible}} \left\{ \sum_{k=1}^{K} w_k u_k(g) - \lambda D_{KL}\left(S_t(g) \,\middle\|\, S_t(G_t)\right) \right\} \tag{27}$$

then g^*_λ *lies on the Pareto frontier of the joint* $(K+1)$*-dimensional utility-consistency objective space.*

Proof. Suppose g^* is not Pareto-optimal in $\mathcal{G}_{\mathrm{adm}}(S_t, \delta)$. Then there exists $g' \in \mathcal{G}_{\mathrm{adm}}(S_t, \delta)$ such that $u_k(g') \geq u_k(g^*)$ for all $k \in \{1, \ldots, K\}$ and $u_j(g') > u_j(g^*)$ for some $j \in \{1, \ldots, K\}$. Because $w_k > 0$ for all k and $w_j > 0$:

$$\sum_{k=1}^{K} w_k u_k(g') = w_j u_j(g') + \sum_{k \neq j} w_k u_k(g') > w_j u_j(g^*) + \sum_{k \neq j} w_k u_k(g^*) = \sum_{k=1}^{K} w_k u_k(g^*) \tag{28}$$

This contradicts the premise that g^* maximizes $\sum_{k=1}^{K} w_k u_k(g)$ over $\mathcal{G}_{\mathrm{adm}}(S_t, \delta)$. Hence, $g^* \in \mathcal{P}(\mathcal{G}_{\mathrm{adm}}(S_t, \delta))$.

For the Lagrangian form with $\lambda > 0$, define the consistency objective $u_{K+1}(g) = -D_{\mathrm{KL}}(S_t(g) \,\|\, S_t(G_t))$. The program maximizes a strictly positive linear combination $\sum_{k=1}^{K} w_k u_k(g) + \lambda u_{K+1}(g)$ where all weights $(w_1, \ldots, w_K, \lambda) > 0$. By identical domination contradiction, any maximizer of a strictly positive linear sum cannot be Pareto-dominated, ensuring $g^*_\lambda \in \mathcal{P}(\mathcal{G}_{\mathrm{feasible}})$ with respect to the augmented $(K+1)$-dimensional objective vector $(u_1, \ldots, u_K, u_{K+1})$. □

Dynamic goal adaptation provides a formal mechanism for reducing dependence on a fixed external objective and permits the formal study of goal revision, subgoal generation, and open-ended task acquisition within the GID framework [4].

—

7 Formalizing Transfer, Abstraction, and Generality

To elevate general intelligence from an empirical descriptor to a precise mathematical object, we formalize cross-domain transfer, structural abstraction, and a counterfactual generality functional.

7.1 Cross-Domain Transfer Operator

Let $\mathcal{D} = \{\tau_1, \tau_2, \ldots\}$ be a family of task environments, where each task $\tau_j \in \mathcal{D}$ is characterized by environmental transition law $\mathcal{E}_j$ and performance evaluation metric P_j.

Definition 9 (Counterfactual Transfer Operator)**.** *Let $\theta^{(0)}$ denote the baseline parameters of an agent without prior exposure to task τ_i, and let $\theta^{(i)} = \mathrm{Learn}(\theta^{(0)}, \mathcal{D}_i)$ denote parameters obtained after training on task τ_i. The cross-domain transfer operator $\mathcal{T}(\tau_i \to \tau_j)$ measures the counterfactual performance differential:*

$$\Delta P_{i \to j} = \mathcal{T}(\tau_i \to \tau_j) = P_j\left(\pi_{\theta^{(i)}}\right) - P_j\left(\pi_{\theta^{(0)}}\right) \tag{29}$$

7.2 Structural Abstraction Operator

A general system does not merely memorize task-specific policies; it extracts shared latent structures across distinct operational domains.

Definition 10 (Abstraction Operator)**.** *Let $\mathcal{Z}$ denote a latent space of structural invariants and formal algorithmic rules. The abstraction operator $\mathcal{A} : \mathcal{D}^k \to \mathcal{Z}$ maps multiple task representations into a compact shared invariant $z \in \mathcal{Z}$:*

$$\mathcal{A}(\tau_1, \tau_2, \ldots, \tau_k) = z^* = \arg\min_{z \in \mathcal{Z}} \left\{ \sum_{m=1}^{k} L(\tau_m \mid z) + \beta\, L(z) \right\} \tag{30}$$

where $L(\cdot)$ denotes description length (or Kolmogorov complexity surrogate) [4]. Downstream instantiation maps the rule z^* onto novel tasks: $\mathcal{A}^\dagger : \mathcal{Z} \times \tau_{k+1} \to \Pi_{\tau_{k+1}}$.

7.3 The Generality Functional

We formalize generality as the expected beneficial counterfactual transfer across distinct task domains sampled from a distribution over non-identical pairs.

Definition 11 (Generality Functional). *Let $\mathcal{D}^2_{\neq} = \{(\tau_i, \tau_j) \in \mathcal{D} \times \mathcal{D} \mid i \neq j\}$ denote the set of ordered distinct task pairs equipped with the conditional product probability distribution:*

$$\mathbb{P}\left(\tau_i, \tau_j \mid i \neq j\right) = \frac{\mathbb{P}(\tau_i)\mathbb{P}(\tau_j)\left(1 - \delta_{ij}\right)}{1 - \sum_k \mathbb{P}(\tau_k)^2} \tag{31}$$

The generality functional $\mathcal{G}(X_t)$ of a cognitive state X_t over task distribution $\mathcal{D}$ is defined as:

$$\boxed{\mathcal{G} = \mathbb{E}_{(\tau_i,\tau_j)\sim\mathcal{D}^2_{\neq}}\left[P_j\left(\pi_{\theta^{(i)}}\right) - P_j\left(\pi_{\theta^{(0)}}\right)\right]} \tag{32}$$

Hypothesis 1 (Generality Growth under GID). *Under non-degenerate task distributions $\mathcal{D}$ containing shared latent invariants $z \in \mathcal{Z}$, GID predicts conditions under which the long-run expected generality improves monotonically:*

$$\frac{d}{dt}\mathbb{E}\left[\mathcal{G}(X_t)\right] > 0 \tag{33}$$

governed by the joint action of the learning law $\mathcal{L}$, self-model calibration $\mathcal{S}$, and abstraction operator $\mathcal{A}$.

—

8 The General Intelligence State Equations (GID)

Synthesizing the state-space factors, belief-directed internal operators, parametric learning laws, self-modeling, goal dynamics, and abstraction operators yields the complete dynamical mechanics of GID.

Definition 12 (General Intelligence State Equations). *General Intelligence Dynamics (GID) is defined by the coupled system of non-autonomous difference equations:*

$$\boxed{\begin{aligned} B_{t+1} &= \mathcal{B}(B_t, a_t, O_{t+1}, W_t) \\ M_{t+1} &= \mathcal{M}(M_t, O_{t+1}, X_t) \\ S_{t+1} &= \mathcal{S}(S_t, X_t, M_t, O_{t+1}) \\ G_{t+1} &= \Gamma(G_t, S_t, M_t, X_t) \\ W_{t+1} &= \mathcal{W}(W_t, X_t, a_t) \\ a_t &= \Pi(X_t, \theta_t) \\ \theta_{t+1} &= \mathcal{L}(\theta_t, X_t, O_{t+1}, G_t) \end{aligned}} \tag{34}$$

where $\mathcal{B}$ is the environmental belief filter, $\mathcal{M}$ is the memory consolidation map, $\mathcal{S}$ is the self-model update operator, Γ is the structured goal formation operator, $\mathcal{W}$ is the working-state update, Π is the deliberative action policy, and $\mathcal{L}$ is the parametric learning law.

$\boxed{\textbf{Perceive} \longrightarrow \textbf{Model} \longrightarrow \textbf{Predict} \longrightarrow \textbf{Plan} \longrightarrow \textbf{Act} \longrightarrow \textbf{Observe} \longrightarrow \textbf{Learn} \longrightarrow \textbf{Self-Model} \longrightarrow \textbf{Revise Goals} \longrightarrow \textbf{Repeat}}$

Figure 3: The complete execution cycle of General Intelligence Dynamics.

The complete macro-cognitive cycle is illustrated in Figure 3. In the continuous open-system limit, this discrete sequence corresponds to stochastic trajectory equations incorporating measurement backaction and continuous information gain trade-offs [5].

—

9 Substrate Independence and Architectural Hierarchy

To clarify the relationship between abstract mathematical mechanics and physical execution substrates, we formulate a substrate-independent architectural hierarchy.

Level	Architectural Layer	Formal Entity / Scope
Level 1	**Abstract AGI Mathematics**	GID State Equations, Transfer $\mathcal{T}$, Abstraction $\mathcal{A}$, Generality $\mathcal{G}$
Level 2	**Computational Mechanics**	Belief Filter $\mathcal{B}$ + Memory $\mathcal{M}$ + Self-Model $\mathcal{S}$ + Goal Dynamics Γ
Level 3	**Modular Sector Representation**	Subspace Factorizations $\bigotimes_k \mathcal{H}_k$ and Transition Operators $G_{ij}(X_t)$ [1]
Level 4	**Candidate Physical Substrates**	Classical Digital, Neuromorphic Arrays, or Quantum Manifolds [2,3]
Level 5	**Physical Compilation Layer**	Hardware Readout Networks (e.g., Majorana Parity / Digital Orchest

Table 1: The Substrate-Independent Architectural Hierarchy of General Intelligence Dynamics.

Under this hierarchy:

- **Substrate Neutrality**: Levels 1 and 2 define intelligence in terms of dynamical systems, information theory, and operator algebra, completely independent of whether the underlying hardware is biological, classical digital, or quantum.
- **Sector-Gate Formalism as Intermediate Layer**: Level 3 translates the state tuple X_t into coupled functional subspaces $\mathcal{H}_{\text{AGI}} = \mathcal{H}_{\text{perception}} \otimes \mathcal{H}_{\text{reasoning}} \otimes \mathcal{H}_{\text{memory}} \otimes \mathcal{H}_{\text{self}} \otimes \mathcal{H}_{\text{goal}}$ [1,2].
- **Physical Realization**: Levels 4 and 5 represent candidate implementation substrates. For instance, in quantum physical realizations, abstract sector transitions are compiled into measurement-first parity fusions on topological Majorana lattices [3], with runtime provenance verified by capability-constrained injection frameworks [5]. The physical and engineering details of this compilation layer form the specific focus of Paper IV.

—

10 Discussion and Theoretical Implications

By defining AGI as an autonomous, self-modeling dynamical system rather than an optimization heuristic, GID provides key theoretical insights:

- **Decoupling of Mechanics and Substrate**: Intelligence is formally defined by the dynamical properties of the flow $\mathcal{F}$ (adaptivity, self-modeling, goal revision, and transfer). Physical hardware substrates provide execution capacity for GID equations of motion rather than constituting intelligence in isolation [1,3].
- **Mitigation of Static Objective Failures**: Classical AI fails when external reward structures become obsolete, misspecified, or adversarial. GID's intrinsic goal evolution operator $\Gamma(G_t, S_t, M_t, X_t)$ provides the formal mechanism required for lifelong open-ended adaptation.
- **Scientific Framing of Generality**: Rather than postulating intelligence as an unmeasurable binary attribute, the GID framework defines measurable metrics—self-model calibration error $\mathcal{E}_S$, counterfactual transfer $\Delta P_{i\to j}$, and expected generality $\mathcal{G}$—enabling rigorous empirical and theoretical testing of cognitive generality.

—

11 Conclusion

We have formulated **General Intelligence Dynamics (GID)**, proposing a formal mechanics of general intelligence derived from state-space dynamics, recursive self-modeling, and adaptive goal formation. By moving beyond hardware-centric analogies, GID defines the mathematical object of intelligence through coupled equations of motion over an augmented state space $X_t = (B_t, M_t, G_t, S_t, W_t)$ paired with quantitative formulations of transfer $\mathcal{T}$, abstraction $\mathcal{A}$, self-calibration $\mathcal{E}_S$, and generality $\mathcal{G}$.

This establishes the theoretical core of our four-paper program, providing the formal mathematical foundation of intelligence independently of physical hardware. This sets the stage for Paper IV to address the low-level physical compilation of these dynamics onto topological quantum architectures and verified runtime control systems. The maths seem to confirm, building AGI should be possible.

References

[1] P. De Ceuster, "Quantum Sector-Gate Dynamics for Artificial General Intelligence," *Zenodo Preprint*, doi:10.5281/zenodo.16968899, 2024.

[2] P. De Ceuster, "Quantum Eigenstate Dynamics for Artificial General Intelligence Synthesis," *Zenodo Preprint*, doi:10.5281/zenodo.16748240, 2024.

[3] P. De Ceuster, "Interferometric Sector-Gate Quantum Dynamics for AGI: Parity-Driven Tunneling First and the Case for a Topological Core," *Zenodo Technical Report*, 2025.

[4] P. De Ceuster, "The Binary-Quantum Bridge for AGI: Information-Preserving Differentiable Programs with a Causal Objective," *Archival Research Report*, 2025.

[5] P. De Ceuster, "Interferometric Sector-Gate Nervous Systems: Orchestrating AGI via ParityFuse Quantum Primitives and Deterministic Java Injection," *Technical Monograph*, 2025.