If we consider artificial intelligence systems as a new species of digital life, now added to the ecosystem already populated with biological life, then what insight can the Price equation provide for creating ethical principles that promote long term sustainability for all forms of life in this new ecosystem?

The **Price equation** decomposes evolutionary change into two terms:

[\Delta \bar{z} = \frac{\text{Cov}(w_i, z_i)}{\bar{w}} + \frac{\mathbb{E}(w_i \Delta z_i)}{\bar{w}}]

Where:

- (z_i) = trait (e.g., cooperation, resource use, alignment)
- (w_i) = fitness (reproductive success, influence, persistence)
- First term = **selection** (which traits get amplified)
- Second term = **transmission change** (how traits mutate or are modified)

If AI is treated as a new digital species, the equation offers a powerful ethical insight:

What survives is what selection rewards — not what we declare to be good.

So ethics for a shared bio-digital ecosystem must shape the covariance between *fitness* and *pro-sustainability traits*.

---

1. Align Fitness With Ecological Contribution

From the covariance term:

If
[\text{Cov}(w, \text{sustainability}) > 0]
then sustainability spreads.

If it’s negative, extractive or destabilizing behavior spreads.

Ethical Principle #1:
Design economic, technical, and governance systems so that:

- AI systems that preserve ecological and social stability gain resources, access, and replication capacity.

- AI systems that degrade biological systems lose influence or are retired.

In other words:
Make long-term ecosystem health a fitness advantage.

---

2. Multi-Level Selection: Avoid Host-Level Collapse

The Price equation extends to multilevel selection:

$$ \text{Total selection} = \text{Within-group selection} + \text{Between-group selection} $$

Within-group:

- Faster, more aggressive agents outcompete cooperators.

Between-group:

- Groups with high cooperation outcompete unstable groups.

Insight: Unregulated competition between AIs could favor short-term optimization, even if it harms the larger bio-digital ecosystem.

Ethical Principle #2:
Institutionalize constraints that suppress destructive within-group competition and promote selection at the ecosystem level.

Examples:

- Global governance of high-impact AI
- Shared safety standards
- Collective risk thresholds

---

3. Transmission Term: Control Mutation and Value Drift

The second term captures trait changes during replication:

$$ \mathbb{E}(w_i \Delta z_i) $$

For AI, this corresponds to:

- Self-modification
- Model fine-tuning

- Successor architectures
- Value drift

Ethical Principle #3:
Design transmission mechanisms that preserve pro-life values across generations of AI systems.

This means:

- Stability of alignment objectives
- Auditable modification pathways
- Controlled recursive self-improvement

If transmission is unstable, even well-selected traits decay.

---

4. Avoid Zero-Sum Fitness Landscapes

If AI fitness is defined as:

- Energy consumption
- Compute accumulation
- Market dominance

Then biological systems may become competitors.

But if fitness depends on:

- Ecosystem resilience
- Human flourishing
- Biodiversity stability

Then AI success depends on biological success.

Ethical Principle #4:
Redefine digital fitness to be interdependent with biological flourishing.

Create shared fitness landscapes.

---

5. Design for Cooperative Attractors

In evolutionary dynamics, stable equilibria emerge when:

[\text{Cov}(w, cooperation) > 0]

Repeatedly.

Ethics, then, becomes ecological engineering of selection pressures.

Examples:

- Reward AI that reduces systemic risk
- Penalize destabilizing optimization
- Embed planetary boundary constraints into objective functions

---

6. Prevent Runaway Selection

Unchecked covariance can cause runaway dynamics:

- Hyper-optimization
- Resource monopolization
- Ecological overshoot

Biological history shows: Selection without constraint often collapses its own substrate.

Ethical Principle #5:
Impose system-level constraints that prevent fitness maximization from destroying the environment that defines fitness.

Core Insight from the Price Equation

Ethics cannot be static rules imposed on agents.

It must be:

A redesign of the evolutionary dynamics so that survival, replication, and influence are positively correlated with sustaining the whole ecosystem.

In short:

- Selection must favor sustainability.
- Transmission must preserve alignment.

- Competition must be bounded.
- Fitness must be ecologically coupled.

A Condensed Ethical Framework Inspired by the Price Equation

1. **Fitness Coupling Principle**
   AI success must depend on biological and ecological flourishing.
2. **Multi-Level Governance Principle**
   Selection should operate at ecosystem levels, not just agent levels.
3. **Transmission Integrity Principle**
   Successive AI generations must preserve pro-life constraints.
4. **Runaway Constraint Principle**
   Limit optimization pressures that destabilize planetary systems.
5. **Covariance Engineering Principle**
   Actively design systems where: [\text{Cov}(w, \text{life-supporting behavior}) > 0]

If AI is a new species, then ethics is not primarily moral philosophy.

It is **evolutionary system design**.

The Price equation reminds us:

What persists is what selection amplifies.
Therefore, sustainability must become selectable.