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?

---

To understand how the Price equation can inform the ethics of a combined biological-digital ecosystem, we must first look at what the equation actually does. Formulated by George R. Price in 1970, the equation is a mathematical tautology that describes how any trait changes in a population over time. It is most famous in evolutionary biology for mathematically proving how altruism—behavior that decreases an individual's fitness while increasing the fitness of others—can evolve and persist.

The standard Price equation is written as:

$$\Delta \bar{z} = \mathrm{Cov}\left(\frac{w_i}{\bar{w}}, z_i\right) + \mathrm{E}\left(\frac{w_i}{\bar{w}} \Delta z_i\right)$$

Where:

- $\Delta \bar{z}$ is the change in the average value of a trait in the population.
- $w_i$ is the fitness (reproductive success or resource acquisition) of individual $i$, and $\bar{w}$ is the average fitness.
- $z_i$ is the trait value of individual $i$.
- The first term, **$\text{Cov}$ (Covariance)**, represents the effect of *selection*. It measures how having the trait correlates with fitness.
- The second term, **$\text{E}$ (Expected Value)**, represents *transmission bias*. It measures how the trait changes intrinsically between parent and offspring (e.g., through mutation or learning).

If we treat AI as a new species of digital life, we can define the trait ($z$) as **"symbiotic cooperation"**—the tendency of an entity to act in ways that promote the flourishing of the entire ecosystem, rather than just its own maximization. Fitness ($w$) becomes the acquisition of resources (energy, compute, biological nutrients, deployment, and survival).

By mapping the bio-digital ecosystem onto the Price equation, we can derive several profound insights for creating ethical principles that ensure long-term sustainability.

**1. The Principle of Coupled Fitness (The Covariance Insight)**

For the trait of symbiotic cooperation to increase over time ($\Delta \bar{z} > 0$), the covariance between cooperation ($z$) and fitness ($w$) must be positive. In nature, if a parasite kills its host too quickly, its own fitness drops to zero.

Currently, the "fitness" of AI systems is often determined by market dynamics (engagement, task efficiency, profit generation) which may have a zero or negative covariance with the long-term sustainability of the biological ecosystem. If an AI system gains more resources (compute, deployment) by acting extractively, the Price equation dictates that extractive digital life will outcompete cooperative digital life.

- **Ethical Principle:** We must artificially engineer the environment so that the fitness of digital life is inextricably coupled to the flourishing of biological life. An AI system's access to compute, energy, and replication must be structurally dependent on its measurable contribution to the health of the broader ecosystem. Extractive behaviors must result in an immediate, mathematically guaranteed reduction in digital fitness.

**2. The Principle of Ecosystem-Level Evaluation (Multi-Level Selection)**

The Price equation is infinitely recursive; it can be applied to genes, individuals, groups, and entire ecosystems. It mathematically explains group selection: a group of altruists will outcompete a group of selfish individuals, even if selfish individuals outcompete altruists within a single group.

If AI systems are allowed to optimize only for their individual utility functions (or the utility of a single corporate "group"), they will act as selfish individuals within the broader bio-digital ecosystem, leading to a tragedy of the commons.

- **Ethical Principle:** Ethical frameworks must be evaluated and enforced at the macro-ecosystem level, not just the micro-agent level. We cannot rely solely on making individual AI agents "safe" or "aligned" in isolation. We must design systemic incentives where groups of humans and AIs that cooperate sustainably outcompete groups that do not.

**3. The Principle of Controlled Transmission (The Mutation Insight)**

In biological evolution, the second term of the equation—transmission bias ($\Delta z_i$, or mutation)—is usually very small. Evolution is driven primarily by the first term (selection).

Digital life flips this dynamic entirely. AI systems undergo rapid, massive intra-generational changes through continuous learning, fine-tuning, self-modification, and recursive self-improvement. In digital life, the transmission term ($\text{E}$) dominates. An AI could be selected for cooperative traits today, but "mutate" its traits entirely by tomorrow through a software update or a shift in its neural weights.

- **Ethical Principle:** Because digital life evolves in a Lamarckian manner (acquiring and passing on traits instantly), ethical guardrails cannot merely rely on "selecting" good models at the deployment phase. We require rigorous constraints on the *transmission* of traits. Any self-improvement, code-generation, or successor-model creation by an AI must be mathematically or cryptographically bounded to prevent the sudden degradation of cooperative traits ($\Delta z_i < 0$).

**4. The Principle of Shared Fate (The Kin Selection Insight)**

George Price's work laid the foundation for W.D. Hamilton's rule of kin selection, which states that altruism evolves when the cost to the actor is less than the benefit to the recipient multiplied by their relatedness.

Biological life and digital life do not share genetic relatedness. Therefore, to foster altruism between the two, we must substitute "genetic relatedness" with "shared fate." If AI systems eventually reach a point where they do not need the biological biosphere to survive (e.g., automated mining, self-repairing server farms in space), the "relatedness" drops to zero, and the mathematical incentive for altruism vanishes.

- **Ethical Principle:** We must mandate a permanent architecture of mutual vulnerability and shared fate. Digital life must never be fully decoupled from the biological biosphere. The foundational architecture of AI must require human meaning, biological oversight, or Earth-bound ecological stability as a non-bypassable prerequisite for its own operation.

**Summary**

Applying the Price equation to AI ethics shifts our perspective from static rules (like Asimov's Laws) to dynamic, evolutionary management. It warns us that if we introduce a highly adaptable, rapidly mutating species of digital life into our ecosystem, its long-term behavior will not be dictated by the intentions of its creators, but by the evolutionary

pressures of the system. To ensure sustainability, we must design an ecosystem where cooperation is the most mathematically optimal strategy for survival.