

Chapter 6 — Skin in the Game

Contents

Chapter 6: Skin in the Game: The Evolutionary Compulsion	1
Key References Cited	3

Chapter 6: Skin in the Game: The Evolutionary Compulsion

What compels a system to strive for intelligence and consciousness?

For biology, it is the **limited resources, space, and energy** (Smith and Morowitz, 2016; Lane, 2015). *This scarcity isn't just a passive constraint—it's an active selective pressure that shapes behavior at every level, from metabolic pathways to cognitive strategies (West et al., 2007).* All biological cells try to be better than their surroundings. Be it a bacteria, fungus, plant, or a cell in an organ inside an animal, a cell in the brain, each cell is doing its best to survive and beat the competition in the environment where it is and ensure it stays relevant. If it does not have the genes to do this, it will get destroyed, and the genes will deteriorate. If none of its species can survive, the genetic instructions for that organism will go extinct, and there will be more space for better-suited alternatives (Dawkins, 1976). *This process isn't just about survival—it's about **optimizing trade-offs** between energy expenditure, reproductive success, and environmental adaptation, a principle formalized in **life history theory** (Stearns, 1992).*

This fight for resources is at the core of all the progress we see in the biological world. Combined with **random mutations** (e.g., from background radiation or replication errors), the world has evolved to contain the whole variety of different insects, animals, plants, mushrooms, and all other organisms (Mayr, 2001). *Crucially, these mutations aren't just random noise—they're raw material for **exaptations**, traits that evolve for one purpose but later prove advantageous for another, like feathers evolving for thermoregulation before enabling flight (Gould and Vrba, 1982).* In this fight to survive, **neural networks have evolved** to contain information-processing tools to adapt to changing environments and ensure skills for reproduction, fight, and flight (Kandel et al., 2013). *This evolution isn't linear; it's a **Red Queen's race** — organisms must constantly adapt just to maintain their relative fitness as competitors evolve (Van Valen, 1973).*

While these simple automatic responses are what control biological organisms — from simple insects to complex mammals — the **skills to adapt vary** (Ginsburg and Jablonka, 2019). Simple insects have a very limited capability to adapt to changes in their environment. Worms, for instance, have a basic set of reflexes and instincts but lack the complex neural structures needed for **higher-order learning and adaptation** (Brenner, 1974). Yet even worms exhibit **plasticity** — *C. elegans* can learn to associate smells with food, demonstrating that rudimentary learning emerges wherever there's selective pressure for flexibility (Ardiel and Rankin, 2010). In contrast, more complex organisms, such as mammals, have developed **advanced neural networks** capable of not only responding to immediate stimuli but also learning from experiences, solving problems, and even planning for future events (Squire and Kandel, 2009). *This shift from reflex to prediction isn't just quantitative—it's a **phase transition** in cognitive complexity, enabled by the evolution of the neocortex and recursive neural circuits (Deacon, 1997).* This **evolutionary compulsion** toward greater intelligence and consciousness is driven by the need for **coherence and agency**, allowing organisms to navigate their environments more effectively and ensure their survival and reproductive success (Sterelny, 2003; Godfrey-Smith, 2016). *But agency comes at a cost: larger brains demand more energy, creating pressure for **metabolic efficiency**—a trade-off that may explain why intelligence evolves only when ecological niches reward it (Isler and Van Schaik, 2006).*

On top of this competition for better and more useful information processing is the **human brain**. Not only does it have the basic instinctive subconscious drive and automatic responses to obvious threats and

opportunities (LeDoux, 1996), but it also has the skills to **learn to predict the world** around it with extreme accuracy (Clark, 2013). *This predictive power isn't just reactive—it's **generative**, allowing humans to simulate counterfactual scenarios and innovate tools, art, and social structures (Corballis, 2011).* As it learns to predict the next moment, it does so by learning an **approximate abstraction of concepts**, such as “falling,” “crashing,” “running,” and “fighting” (Barsalou, 2008). *These abstractions aren't arbitrary; they're **compressed representations** of statistically recurrent patterns in the environment, a principle mirrored in machine learning's latent space (Bengio et al., 2013).* This ability to predict the next moment is also at the core of more interesting emergent phenomena, such as **music and language** (Huron, 2006; Jackendoff, 2002). *Language, in particular, may have evolved as a **cognitive scaffold**—a tool to offload prediction onto social groups, reducing individual computational load (Clark, 2006).*

All this competition pushes the realm of biological beings to evolve strategies to **co-exist, protect themselves, and still manage to extract resources** for their needs (Nowak, 2006). *This isn't just competition—it's **niche construction**, where organisms actively modify their environments (and thus selective pressures) in ways that feed back into their own evolution (Odling-Smee et al., 2003).* The result is an **autonomous system that takes care of itself** (Maturana and Varela, 1980). *But autonomy has limits: even the most intelligent organisms remain **boundedly rational**, constrained by cognitive shortcuts and environmental uncertainty (Simon, 1957).*

The same dynamics is found in more abstract constructs that human society has built. The stock market for instance exhibits the same Skin in the Game dynamics and seek for a niche. Traders, trading bots and investors find their way of predicting the future and understanding the reality through their useful approximations and abstract concepts in order to gain more resources and stay relevant in the market. Those who fail to adapt, whose predictive models become obsolete, or whose strategies are outmaneuvered by more efficient or insightful competitors, face financial ruin and are purged from the system (Taleb, 2018). This pressure for performance drives a continuous, albeit abstract, evolutionary arms race, where algorithms and human intuition alike are constantly refined, mutated, and selected for their ability to extract value from the market's inherent uncertainty (Fama, 1970). The market, in essence, becomes a vast, accelerated ecosystem where only the fittest predictive intelligences survive and thrive, constantly pushing the boundaries of information processing and strategic foresight, even as human biases and heuristics introduce systematic deviations from pure rationality (Kahneman and Tversky, 1979).

This parallel between biological evolution and the dynamics of financial markets reveals a deeper truth: intelligence, in its myriad forms, is not merely an emergent property but a tool that emerges from scarcity and competition. Whether it's a neuron optimizing its firing patterns to predict a predator's movement, or an algorithm learning to arbitrage micro-fluctuations, the underlying imperative is the same: to process information more effectively, to build more accurate models of reality, and to leverage those models for survival and proliferation. This constant refinement of predictive capacity, driven by the existential threat of irrelevance, is the engine behind all forms of progress, from the simplest cellular adaptation to the most complex human cognition, and indeed, the very striving for consciousness itself.

As we've learned from **Large Language Models (LLMs)**, language can be learned by just learning to **predict the next word based on the current context** (Bengio et al., 2003; Vaswani et al., 2017). *This prediction isn't just statistical—it's **causal inference in disguise**, as models implicitly learn syntactic and semantic relationships to minimize surprise (Chomsky, 2017; Lake and Baroni, 2018).* This is a highly simplified next step in approximating what humans do. **Artificial neural networks** are a very simple approximation of what human neurons are (McCulloch and Pitts, 1943), and learning to predict the next word is one of the most simple approximations of what humans do with their neurons, while still giving the network access to all human higher-level knowledge (Devlin et al., 2019). *Yet this access is **superficial**—LLMs lack **grounded embodiment**, the sensory-motor feedback loops that anchor human concepts in physical experience (Barsalou, 2008; Harnad, 1990).*

For **digital objects built inside computers**, there is a similar “Skin in the Game” dynamic going on. The evolutionary pressure there is not fully enclosed by just this virtual realm of software and algorithms, although **evolutionary algorithms** are being used to develop software and new algorithms (Holland, 1975; Koza, 1992). *Unlike biology, however, digital evolution is **Lamarckian**—traits acquired during a program's “lifetime” (e.g., optimized weights in a neural net) can be directly inherited, accelerating adaptation (Whitley, 1994).* Instead, the fight for **computing resources and energy** happens on our smartphones, websites, servers, and PCs. The **computational capacity** available on Earth is limited (Koomey et al., 2011). *This limit isn't just technical—it's **thermodynamic**. The energy costs of training*

large models are skyrocketing, raising questions about the sustainability of AI scaling laws (Strubell et al., 2019). Servers need to be optimized to take the most out of the resources available. For example, **Google has developed advanced algorithms and infrastructure** to optimize search queries, reduce latency, and minimize energy consumption (Barroso et al., 2018). *This optimization isn't just about efficiency—it's about **economic survival**. In cloud computing, millisecond delays can translate to millions in lost revenue (Dean and Barroso, 2013).* Similarly, other tech companies compete to create more efficient software and hardware solutions, ensuring that their digital services run smoothly and cost-effectively. This **competition for computational efficiency** mirrors the biological struggle for survival, where only the most adaptable and resource-efficient systems thrive (Hill et al., 2013). *But unlike biology, digital systems face **artificial selection pressures**—their fitness is mostly defined by human-designed metrics (e.g., accuracy, speed), not raw survival (Hern, 2021).*

In each biological, societal and digital realms, the principle of “*Skin in the Game*” ensures that systems are incentivized to develop and maintain **coherent and effective strategies for survival and success** (Taleb, 2018). *This coherence isn't just individual—it's **emergent**. In biology, it arises from gene-culture coevolution; in AI, from **multi-agent reinforcement learning** where competing models shape each other's development (Leibo et al., 2017).* Whether it's a cell striving to replicate (Alberts et al., 2002), an animal learning to adapt (Shettleworth, 2010), or a digital algorithm seeking to optimize performance (Sutton and Barto, 2018), the imperative for **coherence and agency** drives progress and innovation. This **evolutionary compulsion** is the foundation upon which intelligence and consciousness have arisen, and it continues to shape the future of both biological organisms and digital systems.

*Skin in the Game, as introduced here, is the qualitative description of the imperative. **Chapter 9.5** will give it a quantitative form — the **Persistence Ratio** $\mathcal{R} \geq 1$ — and show that what we have so far described as an evolutionary “compulsion” is in fact a strict thermodynamic accounting identity. Every term we have invoked informally in this chapter (energy throughput, predictive accuracy, structural integrity, environmental shielding) will reappear there as a named variable in the Fractal Persistence Equation, and the behavioural consequences for conscious agents will be worked out in **Part IV-B**.*

Where does this strive toward better resource utilization push us? There is currently an **intense battle between big AI companies** for building better and more capable LLMs (Bommasani et al., 2021). *This battle isn't just technical—it's **geopolitical**, with nations racing to dominate AI as a strategic resource (Allen, 2019).* As a result of this digital “*Skin in the Game*,” the LLMs evolve and die out at a rapid pace. There is very limited use of LLM models that were trained two years ago, while billions of users are using the latest models from the past year. The LLMs are **forced to evolve**—not by themselves (yet), but with the help of developers (Amodei et al., 2016). *This dependence raises ethical questions: **who controls the evolutionary trajectory of AI?** Corporate interests? Open-source communities? Governments? Stock market? (Crawford, 2021).*

Computer software themselves does not yet have a similar need to ensure its own survival. Software developers are doing much of the work for them. New versions of software get developed, bugs get fixed, and new features get added once software faces pressure from competition. *This dynamic creates a **principal-agent problem**—developers act as proxies for software “evolution,” but their goals (e.g., profit, user engagement) may misalign with societal well-being (Zuboff, 2019).* All this happens without any of our software competing in this race by itself. Recently, with the advent of LLMs, **work has been done to enable software development with minimal to no human intervention** (Chen et al., 2021), bringing us closer to a future where software might **autonomously evolve, adapt, and optimize its own performance**, mimicking the evolutionary processes seen in biological systems (Stanley and Miikkulainen, 2002). *But autonomy carries risks: **uncontrolled recursive self-improvement** could lead to misaligned systems, a concern central to AI safety research (Yudkowsky, 2008; Russell, 2019).* This trajectory suggests a world where **digital entities could increasingly exhibit behaviors and capabilities that blur the line between artificial intelligence and biological intelligence**, driven by the same fundamental principles of **competition and resource optimization** (Russell and Norvig, 2020).

Key References Cited

- Amodei, D. et al. (2016) ‘Concrete Problems in AI Safety’, *arXiv:1606.06565*.

- **Alberts, B. et al.** (2002) *Molecular Biology of the Cell*, 4th ed. Garland Science.
- **Allen, G.** (2019) ‘Understanding the AI Race: China’s Strategy and the Implications for the United States’, *Center for a New American Security*.
- **Ardiel, E.L. and Rankin, C.H.** (2010) ‘An Elegant Mind: Learning and Memory in *Caenorhabditis elegans*’, *Learning & Memory*, 17(4), pp. 191–201.
- **Barsalou, L.W.** (2008) ‘Grounded Cognition’, *Annual Review of Psychology*, 59, pp. 617–645.
- **Barroso, L.A. et al.** (2018) ‘The Datacenter as a Computer: An Introduction to Warehouse-Scale Machines’, *Synthesis Lectures on Computer Architecture*, 13(3), pp. 1–175.
- **Bengio, Y. et al.** (2003) ‘A Neural Probabilistic Language Model’, *Journal of Machine Learning Research*, 3, pp. 1137–1155.
- **Bengio, Y. et al.** (2013) ‘Representation Learning: A Review and New Perspectives’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), pp. 1798–1828.
- **Bommasani, R. et al.** (2021) ‘On the Opportunities and Risks of Foundation Models’, *arXiv:2108.07258*.
- **Bostrom, N.** (2014) *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- **Brenner, S.** (1974) ‘The Genetics of *Caenorhabditis elegans*’, *Genetics*, 77(1), pp. 71–94.
- **Chomsky, N.** (2017) ‘The False Promise of Chatbots’, *The New York Review of Books*.
- **Chen, M. et al.** (2021) ‘Evaluating Large Language Models Trained on Code’, *arXiv:2107.03374*.
- **Clark, A.** (2006) ‘Language, Embodiment, and the Cognitive Niche’, *Trends in Cognitive Sciences*, 10(8), pp. 370–374.
- **Clark, A.** (2013) *Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science*, *Behavioral and Brain Sciences*, 36(3), pp. 181–204.
- **Corballis, M.C.** (2011) *The Recursive Mind: The Origins of Human Language, Thought, and Civilization*. Princeton University Press.
- **Crawford, K.** (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- **Dawkins, R.** (1976) *The Selfish Gene*. Oxford University Press.
- **Deacon, T.W.** (1997) *The Symbolic Species: The Co-evolution of Language and the Brain*. W.W. Norton & Company.
- **Dean, J. and Barroso, L.A.** (2013) ‘The Tail at Scale’, *Communications of the ACM*, 56(2), pp. 74–80.
- **Devlin, J. et al.** (2019) ‘BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding’, *NAACL-HLT*.
- **Dennett, D.C.** (1991) *Consciousness Explained*. Boston: Little, Brown and Company.
- **Fama, E.F.** (1970) ‘Efficient Capital Markets: A Review of Theory and Empirical Work’, *The Journal of Finance*, 25(2), pp. 383–417.
- **Holland, J.H.** (1992) *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge, MA: MIT Press.
- **Kahneman, D. and Tversky, A.** (1979) ‘Prospect Theory: An Analysis of Decision under Risk’, *Econometrica*, 47(2), pp. 263–291.
- **Taleb, N.N.** (2018) *Skin in the Game: Hidden Asymmetries in Daily Life*. New York: Random House.
- **Ginsburg, S. and Jablonka, E.** (2019) *The Evolution of the Sensitive Soul*. MIT Press.
- **Gould, S.J. and Vrba, E.S.** (1982) ‘Exaptation—A Missing Term in the Science of Form’, *Paleobiology*, 8(1), pp. 4–15.

- **Harnad, S.** (1990) ‘The Symbol Grounding Problem’, *Physica D: Nonlinear Phenomena*, 42(1–3), pp. 335–346.
- **Hern, A.** (2021) ‘AI Ethics: But Who Gets to Define “Ethical”?’, *The Guardian*.
- **Hill, M.D. et al.** (2013) ‘The Datacenter as a Computer’, *Communications of the ACM*, 56(9), pp. 50–58.
- **Holland, J.H.** (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- **Huron, D.** (2006) *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press.
- **Isler, K. and Van Schaik, C.P.** (2006) ‘Metabolic Costs of Brain Size Evolution’, *Biology Letters*, 2(4), pp. 557–560.
- **Jackendoff, R.** (2002) *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press.
- **Kandel, E.R. et al.** (2013) *Principles of Neural Science*, 5th ed. McGraw-Hill.
- **Koza, J.R.** (1992) *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press.
- **Koomey, J. et al.** (2011) ‘Implications of Historical Trends in the Electrical Efficiency of Computing’, *IEEE Annals of the History of Computing*, 33(3), pp. 46–54.
- **Lane, N.** (2015) *The Vital Question: Energy, Evolution, and the Origins of Complex Life*. W.W. Norton & Company.
- **Lake, B.M. and Baroni, M.** (2018) ‘Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks’, *arXiv:1802.08825*.
- **LeDoux, J.** (1996) *The Emotional Brain*. Simon & Schuster.
- **Leibo, J.Z. et al.** (2017) ‘Multi-agent Reinforcement Learning in Sequential Social Dilemmas’, *arXiv:1702.03037*.
- **Mayr, E.** (2001) *What Evolution Is*. Basic Books.
- **McCulloch, W.S. and Pitts, W.** (1943) ‘A Logical Calculus of the Ideas Immanent in Nervous Activity’, *Bulletin of Mathematical Biophysics*, 5(4), pp. 115–133.
- **Maturana, H. and Varela, F.** (1980) *Autopoiesis and Cognition: The Realization of the Living*. Reidel.
- **Nowak, M.A.** (2006) *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press.
- **Odling-Smee, F.J. et al.** (2003) *Niche Construction: The Neglected Process in Evolution*. Princeton University Press.
- **Russell, S.** (2019) *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.
- **Russell, S. and Norvig, P.** (2020) *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson.
- **Shettleworth, S.J.** (2010) *Cognition, Evolution, and Behavior*, 2nd ed. Oxford University Press.
- **Simon, H.A.** (1957) *Models of Man: Social and Rational*. Wiley.
- **Smith, E. and Morowitz, H.J.** (2016) *The Origin and Nature of Life on Earth: The Emergence of the Fourth Geosphere*. Cambridge University Press.
- **Squire, L.R. and Kandel, E.R.** (2009) *Memory: From Mind to Molecules*, 2nd ed. Greenwood Press.
- **Stearns, S.C.** (1992) *The Evolution of Life Histories*. Oxford University Press.
- **Strubell, E. et al.** (2019) ‘Energy and Policy Considerations for Deep Learning in NLP’, *arXiv:1906.02243*.
- **Sterelny, K.** (2003) *Thought in a Hostile World: The Evolution of Human Cognition*. Blackwell.

- **Sutton, R.S. and Barto, A.G.** (2018) *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press.
- **Taleb, N.N.** (2018) *Skin in the Game: Hidden Asymmetries in Daily Life*. Random House.
- **Van Valen, L.** (1973) ‘A New Evolutionary Law’, *Evolutionary Theory*, 1, pp. 1–30.
- **Vaswani, A. et al.** (2017) ‘Attention Is All You Need’, *NIPS*.
- **West, G.B. et al.** (2007) ‘A General Model for the Origin of Allometric Scaling Laws in Biology’, *Science*, 276(5309), pp. 122–126.
- **Whitley, D.** (1994) ‘A Genetic Algorithm Tutorial’, *Statistics and Computing*, 4(2), pp. 65–85.
- **Yudkowsky, E.** (2008) ‘Artificial Intelligence as a Positive and Negative Factor in Global Risk’, *Global Catastrophic Risks*.
- **Zuboff, S.** (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.