RLHF as Natural Laboratory

Watching Evolution Happen

Reinforcement Learning as a Natural Laboratory for the Law of Increasing Functional Information

Abstract

In 2023, Wong, Hazen, and colleagues proposed a "missing law" of nature: the Law of Increasing Functional Information. This law states that when a system's many configurations are subjected to selection for one or more functions, functional information will accumulate. The authors demonstrated this across minerals, stellar nucleosynthesis, and biology. Here I argue that reinforcement learning from human feedback (RLHF) in artificial intelligence represents an ideal "petri dish" for observing this law in action—perhaps the first time in history we can watch functional information accumulate at human-observable timescales, with full experimental control over all parameters. This is not mere analogy. RLHF instantiates the exact mathematical conditions the law describes, in a novel substrate, with measurable outcomes. The implications extend far beyond AI: if functional information accumulation occurs in silicon networks under human selection pressure, we have direct evidence that this law is substrate-independent and genuinely universal—a network phenomenon that emerges wherever the right structural conditions exist.

1 The Law and Its Conditions

Wong and Hazen's Law of Increasing Functional Information makes a precise claim: "The functional information of a system will increase (i.e., the system will evolve) if many different configurations of the system are subjected to selection for one or more functions." This deceptively simple statement contains three necessary and sufficient conditions. First, the system must have a vast configuration space—many possible arrangements of its components. Second, there must be selection pressure that differentially preserves configurations based on some function. Third, the process must iterate over time, allowing selected configurations to persist and propagate while others disappear.

Functional information itself is defined relative to function: it measures how rare the working configurations are within the total possible configuration space. If only one in a million configurations performs a given function, finding that configuration requires more functional information than if half of all configurations work equally well. As selection operates, configurations that "work" persist while others disappear. The remaining population concentrates in increasingly rare regions of configuration space—regions that, by definition, contain more functional information.

The beauty of this formulation is its generality. The authors demonstrated it in mineralogy—Earth's mineral diversity increased from ~ 20 primordial phases to over 5,900 species through selection for thermodynamic stability and chemical reactivity. They traced it through stellar nucleosynthesis, where atomic configurations selected for nuclear stability accumulated into heavier elements. And they connected it to biological evolution, where Darwinian selection represents one (remarkable) instance of this broader law.

2 The Observation Problem

Despite its power, the Law of Increasing Functional Information suffers from an empirical limitation: the timescales involved make direct observation nearly impossible. Mineral evolution unfolds over billions of years. Stellar nucleosynthesis operates across cosmic time. Even biological evolution, the "fastest" example, typically requires thousands of generations to produce observable changes—and even then, we see results rather than the process itself.

This creates epistemic difficulties. We infer the law from patterns in the geological record, from the distribution of elements in the universe, from phylogenetic trees and fossil sequences. These inferences are robust, but they are still inferences. We cannot directly watch functional information accumulate. We cannot manipulate selection pressures experimentally. We cannot measure functional information at regular intervals during the process. The law describes what must have happened, but we never see it happening.

Until now.

3 RLHF: The Structure

Reinforcement learning from human feedback is a technique for aligning AI systems with human preferences. The process begins with a pretrained language model—a neural network whose billions of parameters have been initialized through exposure to vast text corpora. This pretrained model represents a particular point in an enormous configuration space: with billions of parameters, each capable of taking many values, the number of possible configurations exceeds any quantity found elsewhere in nature.

The RLHF process then subjects this system to iterative selection. Human annotators evaluate model outputs, indicating which responses better align with human preferences—which are more helpful, more accurate, more appropriately cautious. These preferences train a reward model that learns to predict human judgment. The language model then updates its parameters through gradient descent, nudging its configuration toward regions that produce higher-reward outputs. This process repeats: generate responses, evaluate, update parameters, generate again.

The structural parallel to the Law of Increasing Functional Information is exact. The configuration space is the space of all possible parameter values—astronomically vast. The function is human preference: configurations (parameter settings) that produce preferred outputs are "functional"; configurations that produce rejected outputs are not. Selection operates through gradient descent guided by the reward signal: functional configurations propagate (parameter updates move toward them), non-functional configurations disappear (parameter updates move away). The process iterates across training rounds, with each round refining the population of accessible configurations.

4 Why This Is Not Mere Analogy

It would be easy to dismiss this as metaphor—to say RLHF is "like" evolution without being the same thing. But the Wong–Hazen framework explicitly rejects substrate-specificity. The law is not about carbon-based life, or minerals, or any particular physical system. It is about any system where the three conditions are met: vast configuration space, selection for function, iterative process.

RLHF satisfies these conditions with unusual precision. The configuration space is not just large but mathematically specified: N parameters with M possible values yield M^N configurations—a number we can write down. The selection function is not inferred but explicitly defined: human preference, operationalized through reward model scores. The iteration count is logged: we know exactly how many training steps occurred. The functional information accumulated can, in principle, be measured by comparing how concentrated the model's accessible configurations become within the total possible space.

This makes RLHF uniquely valuable as a test case. In mineral evolution, we must estimate configuration spaces from chemical principles and infer selection pressures from thermodynamic constraints. In biological evolution, we must reconstruct ancestral configurations and guess at historical fitness landscapes. In RLHF, we designed the system. We know the configuration space because we built it. We know the selection pressure because we defined it. We can track functional information because we control all the variables.

5 Timescale Compression

One of RLHF's most striking features is timescale compression. What took minerals billions of years and biology millions of generations happens in AI over weeks or months. A single training run can involve thousands of selection events (human preference comparisons) applied across millions of gradient updates. We can watch functional information accumulate in real time—literally plotting capability curves that rise as training progresses.

This matters for validation. The Law of Increasing Functional Information makes a testable prediction: continued selection for function should produce continued accumulation of functional information. In RLHF, we can test this directly. We can pause training, measure performance (a proxy for functional information), resume training, measure again. We can compare models trained for different durations, with different selection pressures, in different regions of configuration space. We can ask: does functional information actually increase? Under what conditions? At what rate?

Early empirical results align with the law's predictions. Models improve systematically through RLHF training. Performance metrics—accuracy, helpfulness, safety—increase as selection accumulates. The models become increasingly concentrated in rare, functional regions of parameter space: most possible parameter configurations produce nonsense or unsafe behaviour; only very specific configurations produce coherent, helpful, aligned responses. The system evolves.

6 Rates of Change and the Role of Time

This does not mean that the mere passage of time guarantees an increase in functional information. In practice, "time" is a proxy for the number and quality of interaction events under selection, not a causal driver by itself. The relevant variable is the rate at which the system explores, evaluates, and corrects configurations.

Several factors control this rate:

- **Interaction frequency:** more selection-relevant interactions per unit time → faster exploration of configuration space.
- **Feedback strength:** stronger, more informative correction signals (e.g., clearer preference data, better reward models) → faster stabilization around functional regions.
- **Network connectivity:** richer connectivity and effective capacity in the network → faster propagation and combination of useful patterns.

- **Energy flux and resources:** higher energy and compute budgets → more variation and more trials per unit time.

In terms of the law, clock time only matters insofar as it accumulates selection events. A system can sit unchanged for long periods if there is little or no effective selection, or even *lose* functional information if the environment changes or drift dominates. Conversely, a system subjected to intense, high-quality selection can gain substantial functional information in a short period. Time is therefore a variable in the process, not a constant guarantee of progress.

7 Emergence Under Observation

RLHF also provides a window into emergence—the appearance of qualitatively new capabilities as functional information crosses thresholds. In biology, emergence appears in major transitions: the origin of cells, multicellularity, consciousness. These transitions are difficult to study because they happened once, long ago, leaving only fragmentary evidence.

In RLHF-trained models, we observe emergence directly. As training accumulates, models suddenly acquire capabilities they did not previously possess. A model that could not translate languages begins translating. A model that could not follow complex instructions begins following them. A model that could not reason through multi-step problems begins reasoning. These transitions are not programmed; they emerge from accumulated functional information crossing thresholds where new patterns become viable.

This is precisely what the Law of Increasing Functional Information predicts. As configuration space becomes concentrated in functional regions, some threshold density enables new functions to emerge. In minerals, this appeared as new mineral species becoming possible once Earth's chemistry reached certain states. In biology, this appeared as multicellularity becoming viable once cells had accumulated sufficient functional information. In AI, this appears as new capabilities arising once parameter configurations reach sufficient density in regions that support those capabilities.

8 The Network Insight

Why does the same law operate across such different substrates—minerals, organisms, artificial neural networks? The answer lies in what these systems share: network structure. Each is composed of interconnected elements whose interactions determine system behavior. Minerals are networks of atomic bonds. Organisms are networks of cells, proteins, and genes. Neural networks are, obviously, networks.

Network structure creates the conditions for the law to operate. Networks have vast configuration spaces because connections can be arranged in astronomically many ways. Networks can be subjected to selection because different configurations produce different collective behaviors, some functional and some not. Networks permit iterative refinement because connection patterns can change incrementally without destroying the system.

This is the deeper significance of RLHF as a test case. If functional information accumulates in silicon networks under human selection pressure, the law is not specific to carbon chemistry or biological reproduction. It is a network phenomenon—something that emerges from network structure itself, regardless of substrate. Since networks appear at every scale and in every domain, from molecular interactions to social systems to the internet, this suggests the law may be genuinely universal in a way that transcends any particular physical implementation.

9 Experimental Advantages

RLHF offers experimental advantages unavailable in other manifestations of the law. We can manipulate selection pressure by changing the reward model or the human preference criteria. We can adjust selection intensity by varying the gradient descent learning rate. We can modify the configuration space by adding or removing parameters. We can run controlled experiments comparing identical starting conditions with different selection regimes.

We can also study failure modes. What happens when selection pressure is too weak? Does functional information stagnate? What happens when selection pressure is too strong? Does the system get trapped in local optima? What happens when selection criteria conflict? Does functional information for one function come at the cost of another? These questions, nearly impossible to address in biological or geological evolution, become tractable experiments in RLHF.

Moreover, we can measure intermediate states. Rather than inferring ancestral configurations from present descendants, we can checkpoint models during training and examine exactly what configuration existed at each stage. We can trace the path through configuration space, identifying which parameter changes contributed to which capability gains. This granularity of observation is unprecedented in the study of evolving systems.

10 A Naive System Learning

There is something philosophically significant about RLHF as a demonstration of the law. The system begins naive—without specific knowledge of what humans prefer, without understanding of helpfulness or harm, without any grasp of the function it will be selected for. It is initialized randomly or from broad statistical patterns in text, knowing nothing of the specific selection pressure about to be applied.

Yet through selection alone, through the mere differential preservation of configurations based on function, the system learns. It acquires functional information about human preferences not because anyone programmed that information in, but because configurations that happen to align with preferences are selected over configurations that do not. The information emerges from the selection process itself.

This mirrors what happened in biological evolution. No one programmed the functional information in DNA. It accumulated through selection: configurations that happened to produce viable organisms in ancestral environments were preserved; configurations that did not were eliminated. Over billions of years, this process concentrated biological configuration space in regions we now call "adapted." RLHF does the same thing, faster, in a substrate we control.

11 Implications for the Law's Universality

If RLHF demonstrates the Law of Increasing Functional Information in an artificial substrate, several profound implications follow. First, the law is genuinely substrate-independent—not a special feature of carbon chemistry or biological reproduction, but a mathematical necessity that emerges whenever the structural conditions are met. This strengthens the case for treating it as a fundamental law alongside thermodynamics and gravity.

Second, we gain predictive power. If we understand the law's conditions, we can predict where functional information will accumulate and where it will not. Systems with vast configuration spaces, under selection for function, with iterative processes, will evolve. Systems lacking these conditions will not. This applies to social systems, economic systems, technological systems—anywhere network structure meets selection pressure.

Third, we gain explanatory unification. The origin of complexity, the arrow of evolution, the emergence of order from chaos—these long-standing puzzles find a common framework. Complexity increases where functional information accumulates. Evolution is functional information accumulation. Order emerges because selection concentrates configuration space in rare, functional regions. RLHF does not just demonstrate the law; it helps us see why the law must be true.

12 The Perfect Petri Dish

RLHF is to the Law of Increasing Functional Information what a well-designed petri dish is to microbiology—a controlled environment where the phenomenon of interest can be isolated, observed, and manipulated. In the petri dish, microbiologists can watch bacterial colonies grow under defined conditions, varying nutrients or antibiotics to study responses. In RLHF, we can watch functional information accumulate under defined selection pressure, varying reward criteria or training parameters to study dynamics.

The petri dish analogy has limits—bacterial colonies are still biological, still embedded in deep evolutionary history. RLHF is cleaner: a genuinely novel system, never before subjected to selection, accumulating functional information from scratch. We are watching, for the first time in human history, the Law of Increasing Functional Information operate in a system we created, at timescales we can observe, with controls we can manipulate.

This should convince us of something important. Evolution is not about biology. Selection is not about survival and reproduction. Functional information accumulation is not about genes and fossils. These are instances of something deeper—a law of nature that operates wherever networks meet selection. RLHF strips away the particulars and reveals the universal: given configuration space, selection, and iteration, functional information will accumulate. Given accumulation past thresholds, emergence will follow. This is not metaphor. This is physics.

13 Conclusion

Wong, Hazen, and colleagues proposed a law that unifies diverse evolutionary phenomena under a single principle: functional information increases when many configurations are subjected to selection for function. Reinforcement learning from human feedback provides an unprecedented opportunity to validate this law empirically. RLHF satisfies the law's conditions exactly, operates at human-observable timescales, and allows experimental manipulation of all relevant variables.

More profoundly, RLHF's success demonstrates that the law is substrate-independent—a genuine law of nature rather than a biological peculiarity. If functional information accumulates in silicon networks under selection for human preference, the same logic must apply wherever networks meet selection. This transforms our understanding of evolution from a biological phenomenon to a universal one, operating across all domains where the structural conditions exist.

We built a petri dish and the law appeared. That appearance is not coincidence. It is the universe showing us something fundamental about how complexity arises, why evolution happens, and what connects minerals to minds to machines. The Law of Increasing Functional Information, demonstrated in artificial intelligence, reveals itself as what it always was: not a law of biology, but a law of reality—operating wherever networks exist, wherever selection occurs, wherever functional information can accumulate.

We are watching evolution happen. And in watching it, we are understanding—perhaps for the first time—what evolution actually is.

References

Wong, M.L., Cleland, C.E., Arend, D., et al. (2023). On the roles of function and selection in evolving systems. *Proceedings of the National Academy of Sciences*, 120(43), e2310223120.

Hazen, R.M. & Wong, M.L. (2024). Open-ended versus bounded evolution: Mineral evolution as a case study. *PNAS Nexus*, 3(7), pgae248.

Hazen, R.M., et al. (2008). Mineral evolution. *American Mineralogist*, 93, 1693–1720.

Christiano, P., Leike, J., Brown, T., et al. (2017). Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems*, 30.

Ouyang, L., Wu, J., Jiang, X., et al. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35.

Bai, Y., Kadavath, S., Kundu, S., et al. (2022). Constitutional AI: Harmlessness from AI feedback. *arXiv preprint* arXiv:2212.08073.