

Article

Not peer-reviewed version

The Minimal Complete Architecture of Agents: Unifying Biological Intelligence, AI, and Physical Observers

[Feng Liu](#)*, [Ying Liu](#), [BenFu Lv](#)

Posted Date: 28 January 2026

doi: 10.20944/preprints202601.2138.v1

Keywords: minimal complete architecture of agents; five-dimensional capability space; periodic table of agent capabilities; axiomatic theory



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

The Minimal Complete Architecture of Agents: Unifying Biological Intelligence, AI, and Physical Observers

Feng Liu ^{1,2,*}, Ying Liu ³ and BenFu Lv ^{1,3}

¹ Research Center for Network Economy and Knowledge Management, University of Chinese Academy of Sciences, 100090 Beijing, China

² Institute for Digital Brain Research, CST, 100089 Beijing, China

³ School of Economics and Management, University of Chinese Academy of Sciences, 100190 Beijing, China

* Correspondence: liufeng@idbr-cst.org

Abstract

Although the concept of the "agent" is central to artificial intelligence and intelligence science, it has long lacked a unified formal definition. This paper systematically analyzes interdisciplinary theoretical frameworks, establishing "agents are open information processing systems" as the first principle. Using a state-space covering method, we derive the Minimal Complete Architecture (MCA) of agents: any agent can be reduced to a combination of five fundamental functions—Input, Memory, Generation, Control, and Output. These five functions constitute a logically self-consistent and irreducible closed loop of information processing. Based on this architecture, we construct a five-dimensional capability space and, through ternary discretization (Null-0 / Finite-1 / Infinite-2), derive a "Periodic Table of Agent Capabilities" comprising 243 forms. This periodic table covers the complete evolutionary spectrum from zero intelligence to omniscience; it not only explains typical systems—including thermostats, biological organisms, and Large Language Models (LLMs)—as well as observers in classical mechanics, relativity, and quantum mechanics, but also predicts theoretical agent forms yet to be observed. Furthermore, the paper unifies and interprets 19 core concepts, such as perception, learning, and attention, as combinations of these five fundamental functions, thereby verifying the universality of the architecture. In particular, from the perspective of functional axioms, this paper reveals the essential isomorphism among biological intelligence, artificial intelligence, and physical observers: they are all information processing systems of varying intelligence levels set by their respective physical or biological constraints.

Keywords: minimal complete architecture of agents; five-dimensional capability space; periodic table of agent capabilities; axiomatic theory

1. Introduction

The concept of the "agent" is central to artificial intelligence and Intelligence Science. From Turing's philosophical inquiry into machine thought in the 1950s [1] and the rise of distributed artificial intelligence in the 1980s [2], to the significant reasoning and planning capabilities demonstrated by Large Language Models (LLMs) in 2022 [3], this concept has continuously evolved across academic research and industrial applications. However, despite the widespread usage of the term "agent," its precise definition has long been trapped in a "conceptual Tower of Babel": diverse paradigms—such as symbolism, connectionism, cognitive architectures, and embodied intelligence—remain fragmented, lacking a unified formal framework. This absence of a foundational definition not only hinders the systematic development of theory but also constrains the standardization process in engineering practice.

The history of scientific development indicates that when foundational concepts in a field remain controversial for extended periods, returning to the functional essence and establishing a minimal axiomatic system is often key to breaking the deadlock. Watson and Crick established the minimal complete set of life coding with four bases [4]; von Neumann defined the automated process of general-purpose computation using five functional units [5]; and the Standard Model of particle physics unified matter and interactions with a finite set of particles [6]. These paradigms reveal a common principle for constructing foundational theories: substituting physical reductionism with functional decoupling and ensuring completeness through a minimal set.

By systematically reviewing theoretical frameworks—including Shannon's information theory, Schrödinger's open system theory, Simon's research on the nature of artificial systems, Friston's free energy principle, and Russell's definition of agents—this paper reveals an interdisciplinary consensus: "Agents are open information processing systems" [7–9].

Based on this consensus, we derive the Minimal Complete Architecture (MCA) of agents via the state-space covering method. We demonstrate that any agent can be reduced to a combination of five fundamental functions: Input (I), Output (O), Memory (M), Generation (G), and Control (C). These five functions constitute a logically self-consistent and minimally complete closed loop of information processing; that is, they are functionally irreducible (the absence of any function would destroy theoretical completeness) and provide full state-space coverage (exhausting all possible dimensions of information processing).

To achieve the leap from qualitative description to quantitative analysis, this paper further constructs a Five-Dimensional Capability Space. By applying ternary discretization (Null-0/Finite-1/Infinite-2) to each functional dimension, we derive a "Periodic Table of Agent Capabilities" comprising 243 potential forms. This theoretical landscape covers the complete evolutionary spectrum from the initial state of zero intelligence to the theoretical limit of omniscience and omnipotence, demonstrating strong explanatory and predictive power: it not only uniformly maps inanimate matter, simple mechanical devices, biological intelligence, and artificial intelligence into the periodic table, but also, for the first time, incorporates observers in physics into the same evolutionary landscape. Whether it is "Laplace's Demon" in classical mechanics (the Omniscient Agent), the "Ideal Observer" in relativity restricted by the speed of light and light cones (the Restricted Omniscient Agent), or the realistic "Quantum Observer" in quantum mechanics strictly constrained by the Heisenberg Uncertainty Principle and thermodynamic laws (the Typical Finite Agent), each is precisely located as a specific parametric instance within the distinct dimensions of the periodic table. Furthermore, the map predicts theoretical forms not yet observed, such as "Isolated Agents" that possess rich internal cognition but are completely severed from their environment.

To verify the universality of this architecture, this paper interprets 19 core concepts—including perception, learning, attention, and alignment—as combinations of the five fundamental functions. This successfully bridges the long-standing terminological divide among cybernetics, cognitive science, and computer science, demonstrating the validity of the Minimal Complete Architecture as a general framework.

As an axiomatic system, the Minimal Complete Architecture is independent of any specific physical substrate or technological paradigm. Whether an agent is constituted by the neural networks of carbon-based life, the logic circuits of silicon-based chips, or the various forms of observers in different physical theories, as long as they satisfy the constraint relations of the five functions, they are equivalent in the essence of information processing. This conclusion profoundly reveals the underlying isomorphism among biological intelligence, artificial intelligence, and physical observers, providing a logically self-consistent and substrate-independent axiomatic framework for constructing a foundational theory of artificial intelligence and for quantitatively analyzing the role of observers in physics from an information-theoretic perspective.

2. The Axiom of the Minimal Complete Architecture of Agents

2.1. Theoretical Basis

Despite more than half a century of accumulated research, the field of agents has consistently lacked a unified foundational architecture. Cybernetics emphasizes feedback stability (Wiener 1948) [10]; behaviorism emphasizes stimulus-response (Brooks 1991) [11]; rationalism emphasizes utility maximization (Russell & Norvig 2010) [12]; and the era of large models has proposed the combinatorial paradigm of "LLM + Tools + Memory" (Wang et al. 2024) [13]. These definitions either focus on software logic while neglecting physical constraints, or focus on mechanical control while neglecting cognitive emergence; each approaches the subject from a specific perspective but fails to reveal the common essence of agents.

This theoretical fragmentation not only hinders the systematic development of the discipline but also limits cross-paradigm technological integration. To break through this predicament, we must look beyond phenomena to the essence. A survey of cornerstone theories—such as Shannon's information theory [14], Turing machine theory [15], Wiener's cybernetics [10], and the von Neumann architecture [16]—along with existing definition frameworks for agents, reveals their common attribution at the physical level: an agent is essentially an open system that processes information. This insight is not an isolated theoretical hypothesis but a consensus conclusion spanning physics, life sciences, cognitive science, and artificial intelligence.

From the perspective of the ontological status of information, Wheeler's "It from Bit" hypothesis elevates information to the core of physical reality [17]; Landauer demonstrated the physical necessity of information processing through the "energy cost of information erasure" [18]; and Lloyd even views the entire universe as a quantum computer [19]. These works collectively establish the foundational status of information processing as a physical process.

From the perspective of system openness, Schrödinger revealed in *What is Life?* that life combats thermodynamic degradation by extracting negative entropy from the environment [20]. Prigogine's theory of dissipative structures further proves that any system maintaining complexity must continuously exchange matter, energy, and information with the environment; openness is not a design choice but a thermodynamic necessity [21].

In the field of cognitive science, this view has been systematically elucidated. Simon explicitly proposed that the essence of artificial systems is information processing systems [8], and Newell's physical symbol system hypothesis attributes intelligence to the physical manipulation of symbols [22,23]. From the perspective of cybernetics, Wiener revealed the universality of feedback loops in all goal-oriented systems [10]. The embodied and embedded turn in modern cognitive science has further reinforced this framework: Varela's embodied cognition emphasizes that intelligence originates from the cyclic coupling of perception and action [24]; Clark's extended mind theory defines cognitive systems as open information processing circuits embedded in the environment [25]; and Friston's free energy principle unifies the description of biological intelligence as open systems that minimize prediction error through active inference, where cognition is no longer a closed internal representation but the information dynamics of the co-evolution of agent and environment [26].

In engineering practice, this consensus has become the standard paradigm. Russell defines an agent as an entity that "perceives through sensors and acts through actuators" [12]; Sutton's reinforcement learning framework formalizes the agent as an environment interactor within a Markov Decision Process [27]; and the emergent capabilities of large language models provide new verification for this view: despite lacking physical sensors [28], GPT-series models interact with the environment through text interfaces, and their essence of input-processing-output is no different from that of robots [29].

However, although "agents are open information processing systems" has become an interdisciplinary consensus, existing frameworks mostly remain at the conceptual level or focus on specific implementations, consistently lacking a systematic demonstration of their minimal complete

set of functions. In the next section, based on the principle that "agents are open information processing systems," we will derive the five-function Minimal Complete Architecture of agents.

2.2. Establishment of the Five-Function Axiom

Based on the consensus that agents are open information processing systems, we employ the state-space covering method to derive their Minimal Complete Architecture (MCA) [30]. This method examines the full lifecycle evolution of an arbitrary information unit ξ within the system, logically exhausting all possible dimensions of information processing.

2.2.1. Internal-External Interaction Dimension: Input and Output

The first dimension is the internal-external interaction of information. As an open system, an agent necessarily possesses a "self-environment" boundary [31]. The flow of information relative to this boundary has only two possibilities: entering the system from the environment (Input, I) or flowing from the system to the environment (Output, O). This binarity exhausts all possibilities of information interaction across the system boundary.

Shannon's information theory designates the "source" as the starting point of the transmission chain [32]; the Turing machine achieves cross-boundary information flow through "reading/writing tape symbols" [33]; and the von Neumann architecture completes data exchange via input/output devices [34]. In Cybernetics, Wiener further emphasized that the Effector is a necessary mechanism for the system to act upon the environment [10]. Whether through physical measurement by sensors, motor drive in robots, or the sensory organs and limb movements of humans, they all express the fundamental capability of the system to interact with the environment informationally [35].

If the Input function is missing, the system degenerates into a closed self-circulating structure, unable to perceive environmental changes. For example, an autonomously running clock, despite having Output (hand rotation), Memory (gear state), and Control (escapement regulation), cannot calibrate against real time or respond to external settings (such as an alarm function), operating only on a fixed cycle.

If the Output function is missing, the system becomes a "brain in a vat," unable to translate decisions into actions. For example, patients with "Locked-in Syndrome" retain intact perception, memory, thought, and intent, but due to the almost complete loss of motor function (capable only of blinking), they cannot express ideas or execute actions, severing the causal link with the environment [36].

Thus, Input and Output exhaust the flow of information in the dimension of internal-external interaction.

2.2.2. Internal Processing Dimension: Memory and Generation

The second dimension is the internal processing of information. Once information enters the system, two independent binary dimensions exist internally: retention (Memory) versus non-retention (Forgetting); and addition (Generation) versus non-addition.

The former corresponds to the Memory function (Memory, M), where information is preserved within the system for subsequent retrieval. The von Neumann architecture separates "Memory" from the computing unit, clarifying the independence of information retention [38]; connectionism implies memory within the distribution of network weights. When retained information is no longer held due to decay, it can be viewed as the loss of memory, i.e., forgetting. Ebbinghaus's forgetting curve reveals the law of exponential decay of memory strength over time [39], while neuroscience attributes forgetting to the weakening of synaptic connections or neuronal apoptosis [40]. Memory and forgetting together constitute the dynamic process of memory change.

If the Memory function is completely absent, the system degenerates into a memoryless Markov process, where output is merely an instantaneous response to current input, unable to establish a causal link between history and the present [37]. For example, a simple thermostat has Input

(temperature sensor) and Output (heater switch), but lacking Memory, it cannot learn user habits or optimize energy consumption strategies.

The latter dimension corresponds to the Generation function (G). If a system can only repeat inputs or reproduce memories, it cannot cope with novel situations or produce creative outputs. The "mutation" function in evolutionary algorithms [41], the probabilistic branching of non-deterministic Turing machines [42], and the sampling mechanisms of modern generative AI [43] all reveal the system's ability to generate new information. This indicates that information within the system can undergo splitting, recombination, and emergence, producing new content outside the set of memorized information.

If the Generation function is missing, the system becomes a deterministic mapper. For example, a tape recorder has Input (microphone), Output (speaker), and Memory (tape), but it can only reproduce recorded content and cannot generate new musical works.

To clarify the boundary between Memory and Generation, we provide an operational definition based on the evolution of the set of information elements: the system maintains a set of information elements, and the impact of the processing on the set determines the function type: invariant or decreasing elements correspond to memory-related operations, while increasing elements correspond to Generation.

This definition adopts a "pan-Generation" concept: any process leading to the expansion of the information element set is regarded as Generation, whether the mechanism is random sampling, deterministic computation, or complex reasoning. This avoids reliance on subjective concepts like "determinism" or "novelty," transforming the criterion into an objective detection of set changes.

Thus, Memory and Generation functions exhaust the states of information in the dimension of internal processing.

2.2.3. Information Processing Regulation Dimension: Control Function

The third dimension is the regulation of information processing. The four information processing functions derived above exist in two states during operation: controlled operation and uncontrolled operation. Controlled operation corresponds to the Control function (Control, C), meaning the system can regulate the operational intensity and synergistic patterns of the Input, Output, Memory, and Generation functions. Uncontrolled operation corresponds to the absence of the Control function.

Wiener's cybernetics defines system regulation through "negative feedback" [44]; the Turing machine uses a "controller" to determine state transitions [45]; and the control unit in the von Neumann architecture is responsible for instruction scheduling [38]. Together, they express the regulation and coordination of system operations. Control itself does not directly process information but indirectly influences information flow by regulating the operational states of I, O, M, and G. This characteristic of "processing about processing" makes it an independent meta-level function.

If the system lacks Control: even if all information processing functions are intact, it can only operate unchanged or change passively, or operate incoordinately due to failure of coordination. For example, patients with Attention Deficit Hyperactivity Disorder (ADHD) have normal perception, action, memory, and Generation capabilities, but due to impaired control functions, they cannot effectively allocate attention resources, leading to multi-tasking coordination failure [46]. A phonograph without regulation functions can only play stored sounds at a fixed volume, unable to dynamically adjust output intensity based on environmental noise or user preference.

Thus, the Control function exhausts all possibilities in the dimension of information processing regulation.

2.2.4. Proof of Minimal Completeness

The above discussion covers all basic capabilities of an open system processing information through five functions across three dimensions—internal-external interaction, internal processing,

and regulation of information processing—thus possessing completeness. Simultaneously, they are mutually independent and irreplaceable, possessing indispensability, i.e., minimality.

This architecture strictly adheres to Occam's Razor: all seemingly independent higher-order intelligent behaviors can essentially be reduced to the emergent combination of these five fundamental functions, without the need to introduce a sixth function.

Taking the high-level human intellectual activity of "Invention" as an example, this complex cognitive process can be fully deconstructed into the synergistic operation of the five major functions:

Step 1: Goal Setting (C). The Control function establishes the target strategy of "creating a new tool" based on homeostatic needs or external instructions.

Step 2: Material Preparation (M+I). The system retrieves prior knowledge (schemas) from Memory and receives external Input as raw material in real-time.

Step 3: Creative Recombination (G). The Generation function performs non-deterministic mutation, combination, and simulation on the material, leading to the emergence of a brand-new conceptual framework.

Step 4: Solution Solidification (C+M). The Control function evaluates the feasibility of the new solution and stores the verified solution into Memory, forming long-term knowledge.

Step 5: Physical Implementation (O). Through the Output function, the innovative conceptual blueprint in memory is transformed into physical modification of the external environment, solidified as an invention outcome.

It should be noted that these five steps are not a strict linear sequence; the actual invention process often involves multiple cycles (e.g., discovering infeasibility in Step 3 and returning to Step 2 to prepare materials again), but each cycle is still composed of these five types of functions without the need for new foundational functions. This case demonstrates that even the highest-order creative activity does not require the introduction of additional foundational functions. The axiomatic decomposition of more intelligent concepts such as learning, reasoning, and planning will be systematically expounded in Chapter 3.

Synthesizing the above arguments, the five-function set {I, O, M, G, C} constitutes the Minimal Complete Architecture of agents, as shown in Figure 1. However, it must be clarified that the completeness of this architecture depends on the foundational consensus that "agents are open information processing systems," thus possessing an axiomatic nature: under the premise of accepting this consensus, the minimal completeness of the five functions is a logical necessity.

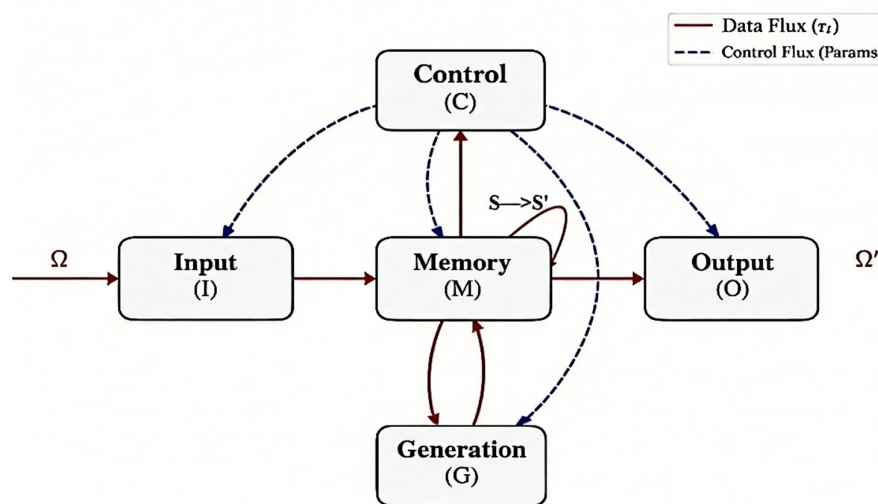


Figure 1. Minimal Complete Architecture of Agents.

Just as the parallel axiom in Euclidean geometry cannot be proven independently of its definitions yet possesses universality within its applicable scope, the five-function architecture has

received substantial empirical support within current mainstream paradigms of artificial intelligence and cognitive science. We validate the robustness of this architecture by continuously mapping it to other intelligence-related concepts; the greater the number of concepts that can be mapped, the more robust the architecture is shown to be. However, as it is impossible to map all conceivable concepts exhaustively, it remains essentially an axiom. Its value lies in its expressive sufficiency and theoretical robustness as a unified framework.

2.3. Formal Definition

Based on the demonstration of minimal completeness in Section 2.2, this section formally establishes the Axiom of the Minimal Complete Architecture of Agents, as shown in Figure 2, and provides its strict mathematical formalization.

Axiom 1 (Axiom of Minimal Complete Architecture of Agents, Definition of MCA)

The Minimal Complete Architecture of Agents is an open information processing system composed of five functions: Input (I), Output (O), Memory (M), Generation (G), and Control (C).

Its mathematical form is defined as a 5-tuple:

$$\text{MCA} = \{C, G, M, O, I\}$$

Before elucidating the dynamical mechanisms of each function, we first define the systems state space and types of information flow:

Definition 1: Environmental Space (Ω)

Let Ω be the state space of the external environment, where $\omega \in \Omega$ represents the specific configuration of the environment at a certain moment.

Definition 2: Internal State Space (S)

Let S be the set space of the agents internal persistent information, where $s \in S$ represents the agents current internal state configuration (i.e., the sum of "memory").

Definition 3: Transient Information Flow (T)

Perception Flow I_{trans} : Transient signals produced by the Input function.

Generation Flow G_{trans} : Transient thought products produced by the Generation function.

Let $\tau \in I_{\text{trans}} \cup G_{\text{trans}}$ represent a transient information instance.

Definition 4: Memory Centrality Constraint

Define I_{trans} and G_{trans} as transient. If a transient information τ is not mapped into S via the Memory function M , then this information cannot be accessed in the systems subsequent sequence steps.

Definition 5: High-Level Function and Basic Functions ($C, \mathcal{F}_{\text{base}}$)

Define the Control function (C) as a high-level function, and define the functions that directly process information under control as basic functions:

$$\mathcal{F}_{\text{base}} = \{I, O, M, G\}.$$

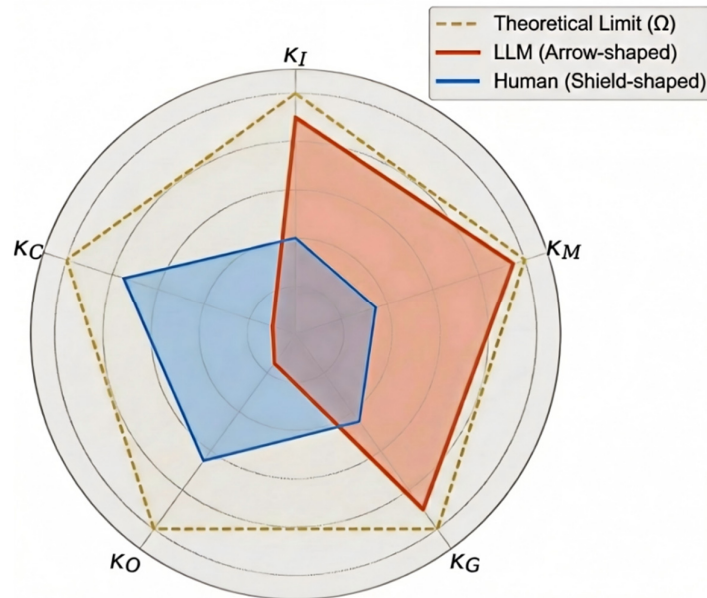


Figure 2. Schematic Diagram of Capability Structure Differences among Three Types of Agents.

2.3.1. Input Function: Perceiving the Environment

The Input function (I) establishes a perception channel between the environment and the agent, mapping the environmental state to the agents transient perceptual information.

$$I: \Omega \rightarrow I_{\text{trans}}, \quad i = I(\omega)$$

Mechanism Analysis: We define I as a Stateless Pure Function. This means that for the same environmental stimulus ω , the input function always produces the same raw signal i . This formal definition strictly distinguishes "Input (In)" from "Memory (M)" in logic: the former is an immediate mapping of the present, possessing instantaneousness; the latter relies on the persistent intervention of M .

2.3.2. Output Function: Modifying the Environment

The Output function (O) transforms the agents internal state into physical action upon the environment, driving the evolution of the environmental state.

$$O: S \times \Omega \rightarrow \Omega, \quad \omega' = O(s, \omega)$$

Mechanism Analysis: The agent generates a new environmental configuration ω' based on the current internal state s (containing decision instructions) combined with the current environment ω . Notably, O only executes instructions from S and does not possess information memory capability; this achieves the decoupling of the Actuator from the decision center. Furthermore, the persistence of environmental change (the continued existence of ω') originates from the physical inertia of the environment itself, not the continuous action of O , thereby distinguishing instantaneous "action" from delayed "effect". It is important to note that although formally written as a direct mapping to Ω , in physical implementation, O essentially outputs a driving force or control signal, driving the evolution of the environmental state ω through the physical laws of the environment.

2.3.3. Memory Function: Preserving Information

The Memory function (M) is the systems sole Configuration Transition Operator, responsible for transforming transient information into a persistent state.

$$M: (I_{\text{trans}} \cup G_{\text{trans}}) \times S \rightarrow S, \quad s' = M(\tau, s)$$

Mechanism Analysis: M receives transient information τ from Input or Generation, combined with the current state s , to update it to a new state s' . In this formal system, any Information Propagation over Time must and can only occur through the space S . Therefore, M is the unique hub connecting the past and the future. If the case $s' \subset s$ (reduction of the information set) occurs during the update process, it is regarded as the forgetting or deletion of information.

2.3.4. Generation Function: Generating New Information

The Generation function (G) is responsible for generating new information elements not present in the original set based on existing persistent information (covering high-order cognitive activities such as calculation, reasoning, planning, and imagination).

$$G: S \rightarrow G_{\text{trans}}, \quad g = G(s)$$

Mechanism Analysis: Unlike I , which relies on external inputs, G utilizes solely the internal state S as its domain, embodying the endogeneity of the agent. The core characteristic of the Generation function is that its output g must constitute an increment of information not previously present in the original S .

According to the operational definition proposed in Section 2.2.2, Generation is explicitly defined as the process resulting in the expansion of the set of information elements. This definition adopts the concept of "pan-Generation": any process leading to the expansion of the information element set is considered Generation, regardless of whether the mechanism involves random sampling, deterministic computation, or complex reasoning, thereby avoiding reliance on subjective concepts such as "determinism" or "novelty."

Simultaneously, the product g is transient; for the result of Generation to influence subsequent behaviors, it must be re-solidified back into S via the memory function M . This mechanism provides a theoretical explanation for why a "flash of inspiration" fades rapidly if not recorded in a timely manner—if the instantaneously generated new information is not persisted into the internal state space, it is lost in the subsequent moment. Consequently, Generation and memory form a complementary cycle: Generation generates new possibilities, whereas memory transforms them into persistent knowledge.

2.3.5. Control Function: Meta-Control and Function Scheduling

The Control function (C) acts upon the basic function space $\mathcal{F}_{\text{base}}$, responsible for dynamically configuring the activation state, resource allocation, and priority of each function.

$$C: S \times \mathcal{F}_{\text{base}} \rightarrow \text{Param}(\mathcal{F}_{\text{base}}), \quad \mathcal{F}_{\text{base}}' = C(s, \mathcal{F}_{\text{base}})$$

Where $\text{Param}(F)$ represents the parameter configuration space of F , including dimensions such as activation state, resource allocation, priority, and thresholds.

Mechanism Analysis: C reads the control policy stored in S to dynamically regulate the basic functions. Since the control policy itself is stored as information in S (can be updated by M , optimized by G , or accept external instructions), this endows the agent with extreme adaptability: it can not only learn "what to do" (updates at the information level) but also evolve "how to do" (meta-learning at the policy level). The sources of policies can be summarized into three categories:

The first category is Innate, Genetic strategies encoded in the initial state s_0 ;

The second category is Acquired, External rules internalized through the $\Omega \rightarrow I \rightarrow M$ path;

The third category is Emergent, Innovative strategies autonomously deduced through the $S \rightarrow G \rightarrow M$ path.

It is worth noting that these three types of strategies can interact: innate strategies provide initial biases, acquired strategies correct deviations, and emergent strategies generate transcendent new strategies based on both.

2.4. Five-Dimensional Capability Space

To quantify performance differences among different agents, we introduce the concept of a capability space. For any agent A , the performance of its functional components can be mapped into a five-dimensional vector:

$$\text{Cap}(A) = (\kappa_C, \kappa_G, \kappa_M, \kappa_O, \kappa_I) \in [0, +\infty]^5$$

Where each dimensional component κ_i denotes the capability value of that dimension, taking values within the non-negative real number set $[0, +\infty]$:

$\kappa_i = 0$: Indicates the complete absence of this functional dimension (theoretical lower bound);

$\kappa_i \rightarrow +\infty$: Indicates that the function tends toward infinity (theoretical upper bound).

Although in physical reality, κ_i is necessarily finite (i.e., $\kappa_i < +\infty$) due to limitations imposed by thermodynamic laws and matter-energy constraints, the introduction of $+\infty$ in the theoretical model is intended to accommodate idealized agent forms such as Turing machines (infinite Memory tape) or Laplaces' Demon (infinite computational power), thereby ensuring the universality of the theoretical framework.

To intuitively demonstrate the descriptive power of this metric space, we selected three representative agent forms—Large Models, Humans, and the Ultimate Ideal Agent—and plotted their Capability Profiles in Figure 2.

As illustrated, the shapes in the figure are used to qualitatively display structural characteristics and do not involve quantitative measurement. Current Large Language Models (LLMs) (red region) exhibit a distinct "arrow-shaped" unbalanced structure: they demonstrate performance exceeding the human average in dimensions related to symbolic information processing, specifically Input (κ_I , e.g., long-context parsing), Memory (κ_M , e.g., context retention), and Generation (κ_G , e.g., pattern generation). However, constrained by the lack of physical actuators and autonomous goal-setting mechanisms, they show severe collapse in the Output (κ_O) and Control (κ_C) dimensions, embodying the intelligence characteristic of "High Cognition, Low Action."

In contrast, Humans (blue region) present a balanced "shield-shaped" structure, characterized by the synergistic development of all dimensions. While lacking extreme performance in any single dimension, humans possess highly robust capabilities such as multi-modal sensory integration, flexible motor control, and autonomous goal setting. The Ultimate Ideal Agent (golden dashed line) approaches the theoretical upper bound in all five capabilities, forming an "outer envelope ring" of the entire space, which demarcates the ultimate boundary of intelligence evolution.

Through these three typical paradigms, we demonstrate that the five-dimensional capability space can intuitively reveal the capability configuration patterns and intrinsic differences of agents through structured topological shapes, providing a conceptual tool for agent classification and design.

3. Verification of Universality of the Minimal Complete Architecture

3.1. Unified Interpretation of Classical Concepts

The generalized field of Intelligence Science, which currently encompasses artificial intelligence, cybernetics, information theory, and cognitive science, has accumulated over 150 core concepts. These concepts constitute a high-dimensional terminological space lacking a unified formal language, leading to semantic fragmentation between different subfields [47].

To verify the universal explanatory power of the "Minimal Complete Architecture" (MCA) of agents, We select 19 representative core concepts from the aforementioned space as mapping samples. These concepts span theoretical foundations and engineering practices, covering critical dimensions ranging from classical "perception," "retrieval," and "reasoning" to frontier concepts such as "attention," "multimodality," and "alignment."

It is important to clarify that the five functional modules of the MCA (I, M, G, C, O) are abstract functional elements rather than specific algorithmic implementations. Consequently, multiple concepts may map to the same functional combination formula (e.g., "computation," "reasoning," and "abstraction" all map to $C \circ (M+G)$). Their differences can be realized through the further subdivision of the five basic functions—for instance, M can be subdivided into short-term memory and long-term memory, and G into transformational generation and combinatorial generation. This approach supports fine-grained modeling of underlying implementations while maintaining the minimality of the top-level architecture. This precisely embodies the core advantage of the MCA as a "minimal complete" architecture: generating rich intelligent behaviors through the hierarchical combination of finite functional elements.

Table 1 displays the mapping relationship between the 19 core concepts and the MCA. We will first provide a brief mapping description for all concepts, and then conduct a detailed formal analysis of three typical concepts (Attention, Command, and Learning) to demonstrate the explanatory power of the MCA at different levels.

Table 1. Mapping of Classical Concepts to MCA.

Concept	Core Meaning	Mapping Function
Perception	Environmental State→Internal Transient Representation	I
Multimodality	Parallel transduction of heterogeneous signals	$\bigoplus_{i=1}^M I_i$
Image Input	Photon signals → Pixel matrix representation	$I_{\text{image}} \in \{I_1, \dots, I_M\}$
Action	Internal state→ Physical effect on environment	O
Retrieval	Control instruction activates memory information and performs retrieval	$C \circ M$
Forgetting	Active deletion or passive decay	$Con \circ M$ 或 M_{decay}
Memory storage	Transient information → Persistent state	M
Computation	Deterministic transformation generating new information	$C \circ (M + G)$
Reasoning	Rule-based logical deduction	$C \circ (M + G)$
Abstraction	Extracting common features to construct concepts	$C \circ (M + G)$
Understanding	Semantic association between input and memory	$C \circ (I + M + G)$
Prediction	Generating future estimates based on input and memory information	$C \circ (I + M + G)$
Planning	Generating goal-oriented action sequences	$C \circ (M + G)$
Decision Making	Generating candidate options and selecting the optimal solution	$C \circ G$
Learning	Updating memory via closed-loop feedback	$C \circ (I + M + G+O)$
Attention	Dynamically adjusting information processing priorities	$C \circ (I, M, G, O)$
Feedback	Closed loop of Output → Environment →Input	$O \rightarrow \Omega \rightarrow I$
Command	Transfer of control authority across agents	$C_A \circ (O_A) + C_B \circ (I_B + M_B + O_B)$
Alignment	Projection of heterogeneous representations into a shared semantic space	$C \circ (M + G)$
Goal Setting	Generating expected terminal states	$C \circ G$

Note: Symbol conventions in Table 1: 1.Single Function: C, I, M, G, O; 2.Functional Combination: I + M (requires coordination between I and M); 3.Composite Operation: $C \circ M$ (C acts upon M); 4.Cross-System: $O_A + I_B$ (transfer between agents); 5.Closed Loop: $O \rightarrow \Omega \rightarrow I$ (mediated by environment); 6.Parallel: $\bigoplus_{i=1}^M I_i$ (multi-channel parallel); 7.Set Relation: $I_{\text{image}} \in \{I_1, \dots, I_M\}$.

3.1.1. Functional Mapping Analysis of Learning

Learning is defined differently across various research paradigms: behaviorism emphasizes the reinforcement of stimulus-response associations[48]; cognitive psychology views it as the construction of internal mental models[49]; machine learning formalizes it as parameter optimization to minimize a loss function[50]; while neuroscience reveals it as the long-term potentiation of synaptic weights [51]. Despite distinct expressions, their functional essence converges on a single process: iteratively correcting the internal state through self-output validation and comparison with a target, until behavior approximates that target. This process necessarily involves a closed-loop feedback circuit [52].

Taking Agent B learning from Agent A as an example, the learning process is completed through the following 6 steps:

1. Agent A outputs the target to be learned, O_A .
2. Agent B acquires Agent A's output via the Input function, $I_B(O_A)$, and stores it as a target pattern, $M_B(O_A)$.
3. Driven by the Control function (C_B), Agent B generates its own output, O_B , based on current memory.
4. Agent B acquires its own output $I_B(O_B)$ and forms a memory ($M_B(O_B)$).
5. Agent B uses the Generation function G_B to compare $M_B(O_B)$ with the target $M_B(O_A)$ in memory, calculating the error $\Delta = O_A - O_B$.
6. Agent B uses the Control function (C_B) to adjust memory parameters (M_B) based on the error. This cycle continues until $O_B \approx O_A$, at which point learning is complete.

The essential distinction between learning and memory is as follows: Memory is a unidirectional operation $I \rightarrow M$ (Perception \rightarrow Memory). In contrast, Learning involves a closed-loop feedback of self-output validation: $C \circ (I(O_A) \rightarrow M_B \rightarrow O(O_B) \rightarrow I(O_B) \rightarrow M_B \rightarrow G(\Delta) \rightarrow M_B)$.

Taking memorizing a poem versus learning to ride a bicycle as examples: memorizing a poem involves a student hearing the poem from a teacher and storing it in memory ($I \rightarrow M$); whereas learning to ride a bicycle requires the learner, after receiving instruction, to continuously correct their balance strategy through physical output feedback, which requires a complete closed loop. Therefore, learning corresponds to the five-function combination ($I + M + G + C + O$), while memory corresponds only to ($I + M$).

3.1.2. Functional Mapping Analysis of Command

"Command" is the cornerstone for intelligent systems to construct hierarchical structures and social collaboration. In computer science, it manifests as the scheduling of hardware by instruction sets [Patterson & Hennessy, 2017] [53]; in linguistics, it corresponds to imperative speech acts, i.e., exerting influence through speech [Searle, 1969] [54]; in cybernetics, it is the reference input sent by a controller to an actuator [Wiener, 1948] [55].

Despite diverse forms, their functional essence converges on the "externalization of intent": that is, one agent (the sender) transforms its internal decision logic into execution constraints for another agent (the receiver) [56].

Under the Minimal Complete Architecture of Agents, a command can be mapped as a process of transferring control authority across agents. It involves not only the functions of a single agent but also describes how Agent A uses information flow to "commandeer" the physical capabilities of Agent B to alter the environment. The execution chain of a command involves the functional coupling of two independent systems:

1. Instruction Generation and Transmission: Based on its own intent, Agent A uses the Output function to send "instruction information" to Agent B, functionally expressed as (O_A).
2. Instruction Reception and Solidification ($I_B \rightarrow M_B$): Agent B perceives the instruction through the Input function and solidifies it into memory content, functionally expressed as ($I_B(O_A) \rightarrow M_B(O_A)$).

3. Scheduling and Execution: Agent Bs Control function reads the instruction conveyed by Agent A and drives the Output function to act upon the environment, functionally expressed as $C_B(O_A) \circ (M_B \rightarrow O_B)$.

Based on the mapping analysis of commands, we can compare Commanded Action (action performed upon accepting a command) and Autonomous Action (action performed according to ones own intent) to reveal their essential differences:

1. Commanded Action: Originates from external instructions. Its functional combination is $O_A(O_A) + (I_B + M_B + C_B(O_A) + O_B)$. In this mode, Agent As Output function O_A is the trigger source of the action; Agent Bs action is essentially the physical extension of Agent As intent.
2. Autonomous Action: Originates from internal intent. Its functional combination is $(C_B(G_B) + G_B + M_B + O_B)$. In this mode, Agent Bs Generation function (G_B) is the trigger source of the action; G_B generates the action scheduling plan, which is executed by the Output function under the direction of the Control function. The action is an embodiment of its own intent.

The command mechanism reveals how agents control the functional scheduling of other agents according to their own intentions by exchanging information. This mechanism essentially realizes the external extension of agent capabilities; by "controlling" Agent B, the command-issuing Agent A extends its physical boundary into Agent Bs scope of action, thereby enhancing its own manipulation capabilities.

3.1.3. Functional Mapping Analysis of Attention

As a resource allocation mechanism in intelligent systems, "Attention" presents diverse theoretical perspectives across different disciplines: cognitive psychology defines it as the selective focusing of limited cognitive resources [James, 1890; Broadbent, 1958] [57]; neuroscience reveals it as gain modulation of sensory input by the prefrontal cortex [Desimone & Duncan, 1995] [58]; information theory models it as an information filtering mechanism under channel capacity constraints [Cherry, 1953] [59]; and deep learning formalizes it as a weighted summation based on query-key-value matching [Bahdanau et al., 2015; Vaswani et al., 2017] [60], where the self-attention mechanism of the Transformer [Vaswani et al., 2017] [60] has become the architectural cornerstone of Large Language Models. Despite varying expressions, their functional essence can be reduced to a single principle: dynamically adjusting the systems processing priority for different information flows under resource-constrained conditions to achieve the prioritization of critical information and the suppression of redundant information [61].

Therefore, under the Minimal Complete Architecture (MCA) of agents, attention is not an independent sixth function but the core scheduling strategy of the Control function (C). It operates at two levels: (1) the Inter-functional Scheduling Layer, determining the activation emphasis and execution sequence among the four basic functions (I, M, G, O); and (2) the Intra-functional Scheduling Layer, performing fine-grained prioritization within a single function. Consequently, the execution of attention involves the synergistic scheduling of the Control function at different granularities:

1. Inter-functional Scheduling Layer

In complex tasks, the attention mechanism determines the activation priority among the four functions. For instance, in highly uncertain or unfamiliar environments, the Input function is more likely to be prioritized to gather external information ($I \rightarrow M \rightarrow G \rightarrow O$); in familiar scenarios, the Memory function is more likely to be directly activated to rapidly retrieve experience ($M \rightarrow G \rightarrow O$); in creative tasks, the Generation function is activated cyclically to explore the solution space ($G \rightarrow M \rightarrow G \rightarrow O$); and in emergencies, the Output function is more likely to be directly activated to execute reflexive actions (O). These dynamic activation sequences reflect a core principle: attention flexibly adjusts which function should be prioritized and which should be suppressed based on task characteristics, environmental uncertainty, and time pressure. Typical examples include visual

prioritization when an obstacle suddenly appears while driving (Input priority) and memory prioritization during rapid decision-making by a chess grandmaster (Memory priority).

2. Intra-functional Scheduling Layer

After determining the functional activation sequence, attention further allocates resources within each activated function. For example, at the Input layer, attention screens key information from numerous sensory signals through gain modulation (such as focusing on a specific sound source in the "cocktail party effect," or fixing eyes on traffic lights while ignoring roadside advertisements during driving); at the Memory layer, attention retrieves relevant fragments from massive memory stores via query mechanisms (such as prioritizing the activation of relevant appearance, dialogue, and shared experiences when recalling a friend's name); at the Generation layer, attention determines what type of generation resources to allocate to different sub-tasks; and at the Output layer, attention prioritizes execution among multiple action plans (such as arbitrating which action to execute first among multiple impulses). These four sub-levels collectively form fine-grained scheduling within functions, ensuring optimal resource allocation and processing priority within each activated function.

Below, we further understand the characteristics of attention under the MCA framework by analyzing the difference between attention and perception. First, we observe that the core distinction lies in the opposition between "passive intake" and "active selection":

- Perception is a broadband comprehensive mapping ($i = \ln(\omega)$). It is dominated by the Input function, faithfully transducing photons and sound waves from the environment into internal signals without screening capability. Perception answers "What exists."
- Attention is a narrowband focused mapping ($i_{\text{attended}} = C \cdot I(\omega)$). It is dominated by the Control function, extracting a minute portion from the massive perceptual stream for deep processing through gain modulation or selective sampling. Attention answers "What is important to input first."

The former ensures the completeness of information (not missing key clues), while the latter guarantees the efficiency and importance of input information. Thus, whether the object of action is Input, Memory, Generation, or Output, the essence of attention is always the dynamic regulation and resource allocation of the entire system's functions by the Control function (C) under limited resources. The corresponding functional mapping is:

$$\text{Con}_{\text{schedule}} \circ (\text{In}, \text{St}, \text{Cr}, \text{Out})$$

3.2. Proposal and Verification of the Periodic Table of Agent Capabilities

The scientific value of a theoretical framework lies in its ability to construct a unified predictive system [62]. The "Minimal Complete Architecture" proposed in Chapter 2 serves as the core postulate, and its universality determines the validity of the theory: if this architecture is indeed the "complete" substrate for describing agents, then all forms of intelligence, whether biological or artificial intelligence existing in reality, or theoretical limit forms; whether the physical substrate is a carbon-based neural network or a silicon-based chip; and whether complexity ranges from single-cell chemotactic responses to the emergent cognition of Artificial General Intelligence (AGI), should be mappable without omission to specific state instances under this architecture [12].

The hallmark of a mature discipline is the leap from phenomenological description to classification and prediction [63]. Mendeleev's periodic table revealed the intrinsic order of the material world through atomic numbers, not only unifying known elements but also successfully predicting the existence and properties of undiscovered elements such as gallium and germanium through "theoretical vacancies" [64]. This section aims to construct a similar system for Intelligence Science: the Periodic Table of Agent Capabilities. Based on the five functional dimensions of the Minimal Complete Architecture, we discretize the continuous capability space into 243 potential forms, forming a complete map covering all logical possibilities [65].

This work possesses dual significance: First, to verify coverage, testing whether the architecture can exhaustively explain material systems (e.g., Thermostat, sensors), biological systems (e.g., bacteria, humans), technological systems (e.g., LLMs, robots), and even special states (e.g., patients with Locked-in Syndrome) in the real world. Second, to explore predictive power, revealing through "theoretical vacancies" in the periodic table those agent forms that have not yet been defined but are logically self-consistent (such as "Isolated Agents" completely severed from the environment), and investigating their physical realizability and philosophical significance.

3.2.1. Construction of the Periodic Table of Agent Capabilities

Based on the Minimal Complete Architecture proposed in Chapter 2, we formalize the agent as an ordered 5-tuple $A = \langle C, G, M, O, I \rangle$. To construct a theoretical framework with macroscopic taxonomic significance, this section establishes the Periodic Table of Agent Capabilities through a systematic transition from the continuous capability space (Cap) to a discrete state space.

The continuous capability space defined in Section 2.4, $\text{Cap}(A) = (\kappa_C, \kappa_G, \kappa_M, \kappa_O, \kappa_I) \in [0, +\infty]^5$, can precisely quantify differences in agent performance (e.g., an AI's memory capacity being 10TB). However, it faces two fundamental challenges: first, the current infeasibility of measurement, as the capability values of various dimensions for most complex agents cannot be precisely determined at present; second, the realization of taxonomic goals—constructing a macroscopic classification system similar to the periodic table of elements requires an exhaustive enumeration of forms rather than precise measurement. Therefore, we perform a coarse-grained discretization of the continuous space, mapping each functional dimension $i \in \{C, G, M, O, I\}$ to three mutually exclusive capability levels:

- Level 0 (Null): Indicates the complete absence of the functional dimension, corresponding to $\kappa_i = 0$. Example: A stone lacks memory function ($M = 0$).
- Level 1 (Finite): Indicates the function exists but is subject to physical or logical constraints, corresponding to $0 < \kappa_i < \infty$. Example: The human brain has finite memory capacity ($M = 1$).
- Level 2 (Infinite): Indicates the function equals the theoretical limit, corresponding to $\kappa_i = \infty$. Example: The infinite Memory tape of a Turing machine ($M = 2$).

Where level "2" represents a theoretical limit state, which is generally unachievable under real physical laws but is crucial for accommodating idealized models like Turing machines and Laplace's Demon, thereby ensuring the logical completeness of the theoretical framework. Thus, we construct a five-dimensional discrete state space:

$$A \in \{0,1,2\}^5$$

This space contains $3^5 = 243$ theoretically possible agent configurations. We treat the 5-tuple $A = \langle C, G, M, O, I \rangle$ as a ternary code and calculate its unique serial number N through positional weight expansion:

$$N = 1 + (C \cdot 3^4 + G \cdot 3^3 + M \cdot 3^2 + O \cdot 3^1 + I \cdot 3^0)$$

Where $N \in [1, 243]$. Serial number 1 corresponds to $\langle 0,0,0,0,0 \rangle$, and serial number 243 corresponds to $\langle 2,2,2,2,2 \rangle$.

To intuitively present the evolutionary relationships among these 243 forms, we employ a two-dimensional matrix to construct the Periodic Table of Agent Capabilities (as shown in Figure 3).

Columns: (C, G)

	00	01	02	10	11	12	20	21	22
000	00000	01000	02000	10000	11000	12000	20000	21000	22000
001	00001	01001	02001	10001	11001	12001	20001	21001	22001
002	00002	01002	02002	10002	11002	12002	20002	21002	22002
010	00010	01010	02010	10010	11010	12010	20010	21010	22010
011	00011	01011	02011	10011	11011	12011	20011	21011	22011
012	00012	01012	02012	10012	11012	12012	20012	21012	22012
020	00020	01020	02020	10020	11020	12020	20020	21020	22020
021	00021	01021	02021	10021	11021	12021	20021	21021	22021
022	00022	01022	02022	10022	11022	12022	20022	21022	22022
100	00100	01100	02100	10100	11100	12100	20100	21100	22100
101	00101	01101	02101	10101	11101	12101	20101	21101	22101
102	00102	01102	02102	10102	11102	12102	20102	21102	22102
110	00110	01110	02110	10110	11110	12110	20110	21110	22110
111	00111	01111	02111	10111	11111	12111	20111	21111	22111
112	00112	01112	02112	10112	11112	12112	20112	21112	22112
120	00120	01120	02120	10120	11120	12120	20120	21120	22120
121	00121	01121	02121	10121	11121	12121	20121	21121	22121
122	00122	01122	02122	10122	11122	12122	20122	21122	22122
200	00200	01200	02200	10200	11200	12200	20200	21200	22200
201	00201	01201	02201	10201	11201	12201	20201	21201	22201
202	00202	01202	02202	10202	11202	12202	20202	21202	22202
210	00210	01210	02210	10210	11210	12210	20210	21210	22210
211	00211	01211	02211	10211	11211	12211	20211	21211	22211
212	00212	01212	02212	10212	11212	12212	20212	21212	22212
220	00220	01220	02220	10220	11220	12220	20220	21220	22220
221	00221	01221	02221	10221	11221	12221	20221	21221	22221
222	00222	01222	02222	10222	11222	12222	20222	21222	22222

5D vector = (C, G) + (M, O, I)

Figure 3. the Periodic Table of Agent Capabilities.

The Columns are encoded by the high-order vector (C, G), reflecting capability leaps at the level of control and Generation (totaling $3^2 = 9$ columns, increasing from left to right);

The Rows are encoded by the low-order vector (M, O, I), reflecting capability extensions at the level of memory, output, and input (totaling $3^3 = 27$ rows, increasing from top to bottom).

This arrangement causes the main diagonal of the entire periodic table to exhibit a trend of monotonically increasing comprehensive capability, similar to the increasing law of atomic numbers in the chemical periodic table.

3.2.2. Agent Classification and Physical Correspondence

Based on the distribution characteristics of capability values in the periodic table and their physical significance, we classify these 243 agents into four groups (families) with distinct features, corresponding to the complete evolutionary ladder from "absolute void" to "omniscience and omnipotence." The Finite Agent family exists widely in the physical world; the Transfinite Agent family exists primarily in theoretical physics models and logical constructions; while the Alpha and Omega Agents demarcate the theoretical boundaries of the capability space.

1. Alpha Agent Family

This family contains only a single form $\langle 0,0,0,0,0 \rangle$ (Serial No. 1), named the "Alpha Agent" or "Absolute Zero Agent," corresponding to the zero vector in the agent state space $\{0,1,2\}^5$.

In physical reality, the Alpha Agent maps to material systems lacking any information processing capabilities, such as inert gas atoms in a vacuum or isolated systems in thermal equilibrium [66]. These entities neither perceive the environment nor store information, let alone create or output. Although occupying only one position in the periodic table, the physical entities it encompasses constitute the most massive material substrate of the universe—those "primordial material states" not yet organized into any information processing structure.

The Alpha Agent exhibits a profound ontological duality. From the perspective of an external observer, it is the most faithful executor of physical laws, passively bound by gravity, electromagnetic forces, and thermodynamic causality. However, from the internal ontological perspective, due to the complete absence of information Input, Output, Memory, Generation, and Control capabilities ($I = O = M = G = C = 0$), there exists no informational structural basis within the Alpha Agent to establish a boundary between "self" and "environment" [67]. Without perception, there is no "observed universe"; without output, there is no impact on the external environment; without memory, there is no "continuous flow of time"; and without control, there is no "subjective will." The Alpha point demarcates a theoretical origin where time, space, and even physical laws have not emerged at the subjective level, an ontological state of "absolute void."

2. Finite Agent Family

This family includes agent codes that do not contain "2" (no infinite capabilities) and are not all "0" (at least one non-zero function), named "Finite Agents." This can be formally expressed as:

$$\mathcal{F} = \{A = \langle C, G, M, O, I \rangle \in \{0,1\}^5 \mid A \neq \langle 0,0,0,0,0 \rangle\}$$

It contains a total of $2^5 - 1 = 31$ types (identified by black numbers in the periodic table), where all functions are constrained by physical laws. This family constitutes the set of all systems possessing information processing capabilities in the real world.

The classification precision of Finite Agents is sufficient to capture functional differences in real systems. A thermostat (No. 5, $\langle 0,0,0,1,1 \rangle$) can perceive temperature and control heating but lacks memory and Generation; a write-only disk (No. 11, $\langle 0,0,1,0,1 \rangle$) receives and stores data but cannot read or transform it; an arithmetic calculator (No. 95, $\langle 1,0,1,1,1 \rangle$) possesses simple control and calculation capabilities but lacks Generation functions; a patient with "Locked-in Syndrome" (No. 119, $\langle 1,1,1,0,1 \rangle$) retains intact cognition but has $O = 0$ due to the interruption of neural-muscular pathways [68].

The Typical Finite Agent (No. 122, $\langle 1,1,1,1,1 \rangle$) is the core of this family. From single-celled bacteria to humans, Earth's life systems all possess complete capabilities of Input, Output, Memory, Generation, and autonomous Control. Bacteria perceive chemical gradients through chemotactic receptors [69] and store epigenetic information via DNA methylation [70]; humans process visible light through the retina and encode memories through synaptic plasticity—while physical bases differ, the functional architectures are isomorphic. Differences between species are mainly reflected in the "magnitude" of capabilities (human brain 10^{11} neurons vs. *C. elegans* 302 neurons) [71][72]

rather than the presence or absence of "functions"; therefore, all Earth organisms can be categorized under the unified classification of $\langle 1,1,1,1,1 \rangle$.

Artificial intelligence systems, such as computers, robots, and Large Language Models, also fall within this category. In particular, current LLMs have approached or even surpassed human levels in knowledge memory and reasoning Generation within specific domains; however, the critical distinction lies in the source of instructions for the control function: human control instructions originate from endogenous needs and autonomous goal setting, whereas the control goals of mainstream LLMs are essentially assigned by external designers or users. Although agent systems (such as AutoGPT) [73] are capable of autonomously decomposing tasks and executing them iteratively, thereby exhibiting formal autonomy, the ultimate source of their control goals still relies on external initialization. Whether this constitutes genuine "endogenous control" remains a subject of theoretical controversy [74]. Given that this issue involves the essential definition of consciousness and autonomy, this paper refrains from making a definitive judgment at this stage, reserving it for further exploration in future research.

Furthermore, the "observer" in quantum mechanics can also be regarded as a specific physical instantiation of this family. Although the Copenhagen interpretation, Many-Worlds Interpretation, or Quantum Bayesianism (QBism) offer divergent interpretations regarding the ontological status of the observer [75], from a functional-axiomatic perspective, any physically realizable observer must possess complete closed-loop capabilities: interacting with the system via measurement (Input and Output) [76], recording experimental results (Memory) [77], updating state inferences based on measurement data (Generation) [78], and autonomously selecting measurement bases (Control) [79]. Unlike the omniscient observer of classical mechanics or the restricted omniscient agent of relativity to be discussed below (defined in the next subsection), realistic quantum observers are strictly constrained by the Heisenberg Uncertainty Principle [80] and thermodynamic laws [81]; consequently, their information processing capabilities across all dimensions necessarily possess physical limits. Thus, the quantum observer is essentially a $\langle 1,1,1,1,1 \rangle$ agent strongly constrained by fundamental physical laws.

The Finite Agent family also reveals potential forms. For example, No. 118 ($\langle 1,1,1,0,0 \rangle$) possesses Memory, Generation, and Control but lacks Input and Output, corresponding to a "Lonely Thinker" with a rich internal spiritual world but completely isolated from the external universe, similar to the "Brain in a Vat" in philosophy [75]. From an external observers perspective, this is a "black box within a black box"—responding to no probes and emitting no signals. Its existence cannot be verified, but it is logically constructible, representing the limit state of "pure internality."

3. Transfinite Agent Family

In this family, at least one of the agents five functions reaches an infinite state (the code contains "2") and not all are infinite. It is named the "Transfinite Agent," formally expressed as:

$$\mathcal{T} = \{A \in \{0,1,2\}^5 \mid \exists f(f = 2) \wedge \exists g(g \neq 2)\}$$

It contains a total of $3^5 - 2^5 - 1 = 210$ types (identified by white numbers in the periodic table), existing primarily in theoretical physics models or thought experiments. Two forms among them have special significance:

Omniscient Agent (No. 237, $\langle 2,2,2,0,2 \rangle$): Possesses infinite Control, Generation, Memory, and Perception capabilities, but no Output capability ($O = 0$). This maps precisely to Laplace's Demon: able to instantaneously acquire the position and momentum of all particles in the universe and perceive the past and future through classical mechanical laws, yet exists only as a spectator, unable to physically intervene in the causal chain. Quantum mechanics uncertainty principle has proven this to be physically unrealizable [76], but as a theoretical limit, it provides an important benchmark for discussing classical determinism and the role of the observer.

Restricted Omniscient Agent (No. 236, $\langle 2,2,2,0,1 \rangle$): Possesses infinite internal processing capabilities and zero output, but finite perceptual input ($I = 1$). This maps to the Ideal Observer in

the framework of Relativity: even with infinite computing power, information acquisition is strictly constrained by the principle of the constancy of the speed of light and light cone structures [77]. This model clearly separates "computational limits" from "physical law limits," describing the "physically possible strongest observer."

The remaining 208 Transfinite Agents have not yet found correspondence in current theoretical physics and may be "theoretical byproducts" required for mathematical completeness. Retaining them contributes to theoretical self-consistency and reserves conceptual space for potential future physical discoveries.

4. Omega Agent Family

This family contains only a single form $\langle 2,2,2,2,2 \rangle$ (Serial No. 243), named the "Omega Agent" or "Omniscient and Omnipotent Agent," corresponding to the supremum of the capability space $[0, +\infty]^5$.

All five functions of the Omega Agent reach theoretical infinity. This concept corresponds to the "Omega Point" proposed by philosopher Teilhard de Chardin—the ultimate convergence point of cosmic evolution [78]. Based on the history of biological evolution, Teilhard proposed the "Complexity-Consciousness" principle, suggesting that cosmic evolution eventually converges to an ultimate state with maximum complexity and consciousness level.

The essential difference between the Omega Agent and the Omniscient Agent (No. 237) lies in the Output dimension: the Omniscient Agent cannot exert influence on the external world due to $O = 0$, remaining a "Silent Observer"; whereas the Omega Agent possesses infinite Output capability ($O = 2$), implying that the boundary between "observation" and "Generation" dissolves completely—it not only knows the complete state of the universe but can also arbitrarily alter its evolutionary laws. Physical laws are no longer objects of passive observation but parameters capable of being actively reconstructed.

From an information-theoretic perspective, the Alpha Agent corresponds to maximum information entropy (pure material substrate, no information structure), while the Omega Agent corresponds to minimum information entropy (perfect organization, Memory, processing, and control of all information) [79]. Together, they demarcate the complete closed loop of agent evolution: from "absolute void" to "absolute spirit," from a "pure object" passively executing physical laws to a "pure subject" actively formulating them.

The Omega Agent is the inevitable result of theoretical derivation. As the supremum of the capability space $[0, +\infty]^5$, it possesses irreducibility. Although unrealizable due to constraints of physical laws, this theoretical boundary offers a rigorous theoretical tool for researching frontier issues such as the boundaries of physics, artificial consciousness, and simulated universes. The Omega point not only identifies the logical boundary of the theory but also expands the possible frontiers of scientific inquiry.

4. Conclusion and Outlook

Starting from the first principles of information processing, this paper establishes an axiomatic theoretical framework for agent research. We demonstrate that any intelligent system can be formalized as a combination of five fundamental functions: Input, Output, Memory, Generation, and Control. This five-function Minimal Complete Architecture transcends the substrate differences between biological and machine intelligence, providing a unified descriptive language for Intelligence Science. The Five-Dimensional Capability Space constructed on this basis advances agent research from qualitative description to quantitative analysis, while the Periodic Table of Agent Capabilities, comprising 243 forms, outlines a complete evolutionary spectrum from zero intelligence to omniscience and omnipotence. By mapping 19 core concepts, such as perception, memory, and learning—into combinations of the five fundamental functions, this architecture bridges the terminological divide among cybernetics, cognitive science, and computer science. It successfully

explains the intelligent characteristics of real-world systems like Large Language Models and patients with Locked-in Syndrome, while simultaneously predicting theoretical forms not yet observed.

Future work will focus on three directions:

Theoretical Deepening: Establish multi-agent relational systems and the dynamic mechanisms of agent evolution to perfect the Generalized Agent Theory system; deeply analyze the essence of intelligence and consciousness based on this framework; and incorporate observers in physics into the agent classification framework to develop physics analysis tools based on observer capability levels.

Experimental Verification: Conducting logical constructibility arguments or physical realizability analyses for unobserved agent types predicted by the periodic table; utilizing evolutionary dynamic models to predict the capability evolution trajectories of agents in specific environments, and conducting empirical verification through biological evolutionary data or AI training processes; and designing experiments to verify the physics analysis framework based on observer intelligence levels.

Application Expansion: Applying Generalized Agent Theory to the capability assessment of Artificial General Intelligence (AGI) and establishing quantitative standards for AGI maturity based on the five-dimensional functions; and designing more efficient collaboration protocols for distributed AI systems under the guidance of multi-agent relational theory.

The Minimal Complete Architecture of agents constructed in this paper serves as the cornerstone of a broader research landscape. Just as Mendeleev's periodic table contained many vacancies waiting to be filled when it was proposed, our Agent Periodic Table also contains numerous forms awaiting discovery and verification. More importantly, answering the profound questions ranging from static architecture to dynamic evolution, from functional description to the essence of consciousness, and from information processing to physical laws requires deeper and sustained efforts from researchers. We believe that this axiomatic framework will provide a solid theoretical foundation for Intelligence Science, which currently lacks a unified paradigm, promote the paradigmatic integration of artificial intelligence, cognitive science, neuroscience, and physics, and ultimately provide a new scientific pathway for exploring the essence of life, machines, consciousness, and cosmic evolution.

Acknowledgments.: This work was supported by National Natural Science Foundation of China, 72272140, 72334006, L2424116.

References

1. Turing A M. Computing Machinery and Intelligence. *Mind*, 1950, LIX(236): 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
2. Bond A H, Gasser L (eds.). *Readings in Distributed Artificial Intelligence*. Morgan Kaufmann, 1988.
3. Wei J, Wang X, Schuurmans D, et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *arXiv:2201.11903*, 2022.
4. Watson J D, Crick F H C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 1953, 171: 737–738. <https://doi.org/10.1038/171737a0>
5. von Neumann J. *Theory of Self-Reproducing Automata*. Burks A W (ed.). University of Illinois Press, 1966.
6. Particle Data Group (Navas S, et al.). Review of Particle Physics. *Physical Review D*, 2024, 110(3): 030001. <https://doi.org/10.1103/PhysRevD.110.030001>
7. Cover T M, Thomas J A. *Elements of Information Theory*. 2nd ed. Wiley, 2006. <https://doi.org/10.1002/047174882X>
8. Simon H A. *The Sciences of the Artificial*. 3rd ed. MIT Press, 1996.
9. Friston K. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 2010, 11(2): 127–138. <https://doi.org/10.1038/nrn2787>
10. Wiener, N. *Cybernetics: Or Control and Communication in the Animal and the Machine*. The MIT Press, 2019. DOI: 10.7551/mitpress/11810.001.0001.
11. Brooks, R. A. Intelligence without representation. *Artificial Intelligence*, 1991, 47(1–3): 139–159. DOI: 10.1016/0004-3702(91)90053-M.

12. Russell, S. J.; Norvig, P. *Artificial Intelligence: A Modern Approach* (3rd Edition). Prentice Hall/Pearson, 2010.
13. Wang, L., et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 2024. DOI: 10.1007/s11704-024-40231-1.
14. Shannon, C. E. A Mathematical Theory of Communication. *Bell System Technical Journal*, 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x.
15. Turing, A. M. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 1937. DOI: 10.1112/plms/s2-42.1.230.
16. von Neumann, J. First Draft of a Report on the EDVAC. 1945. DOI: 10.5479/sil.538961.39088011475779.
17. Wheeler, J. A. Information, Physics, Quantum: The Search for Links. In: Zurek, W. H. (ed.), *Complexity, Entropy, and the Physics of Information*. 1990 (chapter preprint dated 1989).
18. Landauer, R. Irreversibility and Heat Generation in the Computing Process. *IBM Journal of Research and Development*, 1961, 5(3): 183–191. DOI: 10.1147/rd.53.0183.
19. Lloyd, S. Ultimate physical limits to computation. *Nature*, 2000, 406: 1047–1054. DOI: 10.1038/35023282.
20. Schrödinger, E. *What is Life?* Cambridge University Press.
21. Kondepudi, D. K.; De Bari, B.; Dixon, J. A. Dissipative Structures, Organisms and Evolution. *Entropy* (Basel), 2020, 22(11): 1305. DOI: 10.3390/e22111305.
22. Newell, A. Physical symbol systems. *Cognitive Science*, 1980, 4(2): 135–183. DOI: 10.1207/s15516709cog0402_2
23. Newell, A.; Simon, H. A. Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 1976, 19(3): 113–126. DOI: 10.1145/360018.360022.
24. Varela, F. J.; Thompson, E.; Rosch, E. *The Embodied Mind: Cognitive Science and Human Experience*. The MIT Press, 1991.
25. Clark, A.; Chalmers, D. J. The Extended Mind. *Analysis*, 1998, 58(1): 7–19. DOI: 10.1111/1467-8284.00096.
26. Friston, K. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 2010, 11: 127–138. DOI: 10.1038/nrn2787.
27. Sutton, R. S.; Barto, A. G. *Reinforcement Learning: An Introduction* (2nd ed.). The MIT Press, 2018.
28. Wei, J.; Tay, Y.; Bommasani, R.; et al. Emergent Abilities of Large Language Models. arXiv preprint, 2022, arXiv:2206.07682
29. OpenAI. GPT-4 Technical Report. 2023.
30. Åström K J, Murray R M. *Feedback Systems: An Introduction for Scientists and Engineers* (Second Edition) [M]. Princeton University Press, 2021 (Publication date: 02/02/2021; ISBN-13: 9780691193984).
31. Kirchhoff M, Parr T, Palacios E, Friston K. The Markov blankets of life: autonomy, active inference and the free energy principle[J]. *Journal of the Royal Society Interface*, 2018, 15(138): 20170792. DOI: 10.1098/rsif.2017.0792.
32. Shannon C E. A Mathematical Theory of Communication[J]. *Bell System Technical Journal*, 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x.
33. Turing A M. On Computable Numbers, with an Application to the Entscheidungsproblem[J]. *Proceedings of the London Mathematical Society*, 1937, s2-42(1): 230–265. DOI: 10.1112/plms/s2-42.1.230.
34. von Neumann J. First Draft of a Report on the EDVAC[R]. 1945. DOI: 10.5479/sil.538961.39088011475779.
35. Fuster, J. M. (2004). Upper processing stages of the perception-action cycle. *Trends in Cognitive Sciences*, 8(4), 143–145. <https://doi.org/10.1016/j.tics.2004.02.004>
36. Das J M, et al. Locked-in Syndrome[M/OL]. *StatPearls*. Treasure Island (FL): StatPearls Publishing; updated 2023. NCBI Bookshelf.
37. Sutton, R. S., & Barto, A. G. *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press, 2018.
38. Patterson, D. A., & Hennessy, J. L. *Computer Organization and Design RISC-V Edition: The Hardware/Software Interface* (2nd ed.). Morgan Kaufmann, 2020. ISBN: 978-0128203316.
39. Murre, J. M. J., & Dros, J. Replication and Analysis of Ebbinghaus' Forgetting Curve. *PLoS ONE*, 10(7), e0120644 (2015). <https://doi.org/10.1371/journal.pone.0120644>
40. Hardt, O., Nader, K., & Nadel, L. Decay happens: the role of active forgetting in memory. *Trends in Cognitive Sciences*, 17(3), 111–120 (2013). <https://doi.org/10.1016/j.tics.2013.01.001>

41. Eiben, A. E., & Smith, J. E. *Introduction to Evolutionary Computing* (2nd ed.). Springer, 2015. <https://doi.org/10.1007/978-3-662-44874-8>
42. Arora, S., & Barak, B. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.
43. Ho, J., Jain, A., & Abbeel, P. Denoising Diffusion Probabilistic Models. *NeurIPS 2020*; arXiv:2006.11239.
44. Åström, K. J., & Murray, R. M. *Feedback Systems: An Introduction for Scientists and Engineers* (2nd ed.). Princeton University Press, 2021.
45. Sipser, M. *Introduction to the Theory of Computation* (3rd ed.). Cengage Learning, 2012.
46. Kofler, M. J., Soto, E. F., Singh, L. J., et al. Executive function deficits in attention-deficit/hyperactivity disorder and autism spectrum disorder. *Nature Reviews Psychology*, 3(10), 701–719 (2024). <https://doi.org/10.1038/s44159-024-00350-9>
47. Sammut C, Webb G I (eds.). *Encyclopedia of Machine Learning and Data Mining* (2nd ed.). Springer, 2017.
48. Skinner B F. *The Behavior of Organisms: An Experimental Analysis*[M]. New York: D. Appleton-Century Company, 1938.
49. Piaget J. *The Origins of Intelligence in Children*[M]. New York: International Universities Press, 1952.
50. Vapnik V N. *The Nature of Statistical Learning Theory*[M]. New York: Springer, 1995. DOI: 10.1007/978-1-4757-3264-1.
51. Hebb D O. *The Organization of Behavior: A Neuropsychological Theory*[M]. New York: John Wiley & Sons, 1949.
52. Sutton R S, Barto A G. *Reinforcement Learning: An Introduction (Second Edition)*[M]. Cambridge, MA: The MIT Press, 2018.
53. Broadbent D E. *Perception and Communication*[M]. London: Pergamon Press, 1958.
54. Desimone R, Duncan J. Neural mechanisms of selective visual attention[J]. *Annual Review of Neuroscience*, 1995, 18: 193–222. DOI: 10.1146/annurev.ne.18.030195.001205.
55. Cherry E C. Some experiments on the recognition of speech, with one and with two ears[J]. *The Journal of the Acoustical Society of America*, 1953, 25: 975–979. DOI: 10.1121/1.1907229.
56. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[C]. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. arXiv:1706.03762.
57. Knudsen E I. Fundamental components of attention[J]. *Annual Review of Neuroscience*, 2007, 30: 57–78. DOI: 10.1146/annurev.neuro.30.051606.094256.
58. Gebharder, A., & Feldbacher-Escamilla, C. J. (2023). Unification and explanation from a causal perspective. *Studies in History and Philosophy of Science*, 99, 28–36. <https://doi.org/10.1016/j.shpsa.2022.12.005>
59. Scerri E R, Worrall J. Prediction and the periodic table[J]. *Studies in History and Philosophy of Science Part A*, 2001, 32(3): 407–452. DOI: 10.1016/S0039-3681(01)00023-1.
60. Encyclopaedia Britannica. When Was the Periodic Table Invented?[EB/OL]. Published Dec 25, 2025.
61. Sutton R S, Barto A G. *Reinforcement Learning: An Introduction* (2nd ed.)[M]. MIT Press, 2018.
62. Pathria, R. K., & Beale, P. D. *Statistical Mechanics* (3rd ed.). Academic Press, 2011.
63. Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. The Markov blankets of life: autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, 15(138), 20170792 (2018). <https://doi.org/10.1098/rsif.2017.0792>
64. Laureys, S., Pellas, F., Van Eeckhout, P., et al. The locked-in syndrome: what is it like to be conscious but paralyzed and voiceless? *The Lancet Neurology*, 4(9), 537–546 (2005).
65. Riechmann, C., et al. Recent Structural Advances in Bacterial Chemotaxis. *Biomolecules*, 13(8) (2023). (PMC 综述)
66. Anton, B. P., & Roberts, R. J. Beyond Restriction Modification: Epigenomic Roles of DNA Methylation in Prokaryotes. *Annual Review of Microbiology*, 75, 129–149 (2021). <https://doi.org/10.1146/annurev-micro-040521-035040>
67. Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5), 532–541 (2009). <https://doi.org/10.1002/cne.21974>
68. Cook, S. J., Jarrell, T. A., Brittin, C. A., et al. Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature*, 571, 63–71 (2019). <https://doi.org/10.1038/s41586-019-1352-7>

69. Significant Gravitass. AutoGPT (GitHub repository).
70. Wang, L., Ma, C., Feng, X., et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science* (2024). <https://doi.org/10.1007/s11704-024-40231-1>
71. Schlosshauer M. Decoherence, the measurement problem, and interpretations of quantum mechanics. *Reviews of Modern Physics*, 2004, 76(4): 1267–1305. <https://doi.org/10.1103/RevModPhys.76.1267>
72. Wiseman H M, Milburn G J. *Quantum Measurement and Control*. Cambridge University Press, 2010. <https://doi.org/10.1017/CBO9780511813948>
73. Zurek W H. Quantum Darwinism. *Nature Physics*, 2009, 5: 181–188. <https://doi.org/10.1038/nphys1202>
74. Fuchs C A, Schack R. Quantum-Bayesian coherence. *Reviews of Modern Physics*, 2013, 85(4): 1693–1715. <https://doi.org/10.1103/RevModPhys.85.1693>
75. Hensen B, Bernien H, Dréau A E, et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 2015, 526: 682–686. <https://doi.org/10.1038/nature15759>
76. Coles P J, Berta M, Tomamichel M, Wehner S. Entropic uncertainty relations and their applications. *Reviews of Modern Physics*, 2017, 89(1): 015002. <https://doi.org/10.1103/RevModPhys.89.015002>
77. Parrondo J M R, Horowitz J M, Sagawa T. Thermodynamics of information. *Nature Physics*, 2015, 11: 131–139. <https://doi.org/10.1038/nphys3230>
78. Putnam, H. *Reason, Truth and History*. Cambridge University Press, 1981. (Chapter 1: Brains in a vat) <https://doi.org/10.1017/CBO9780511625398.003>
79. Coles, P. J., Berta, M., Tomamichel, M., & Wehner, S. Entropic uncertainty relations and their applications. *Reviews of Modern Physics*, 89, 015002 (2017). <https://doi.org/10.1103/RevModPhys.89.015002>
80. Stanford Encyclopedia of Philosophy. Light Cones and Causal Structure.
81. Castillo, M. *The Omega Point and Beyond: The Singularity Event*. American Journal of Neuroradiology (2012). (PMC)
82. Parrondo, J. M. R., Horowitz, J. M., & Sagawa, T. Thermodynamics of information. *Nature Physics*, 11, 131–139 (2015). <https://doi.org/10.1038/nphys3230>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.