

Theory of Functional Agency: Redefining Agency via Operational System Dynamics

Joyer*

* Independent Researcher (ORCID: <https://orcid.org/0009-0007-7246-5092>), Corresponding author: joyer@joyer.jp

Abstract

This paper redefines the concept of "Agency"—long conflated with psychological attributes like "consciousness" or "intention," or specific biological structures—from a functional perspective, formulating a universal axiomatic framework of "Functional Agency." We frame agency not as a static property, but as an "operational process" interacting with the external environment, proposing three axioms (Boundary Maintenance, Internally Initiated Discriminative Connection, and Irreversible State Change) as necessary and sufficient conditions independent of hardware implementation or internal intelligence. By introducing this axiomatic framework, "Functional Agency" is explicitly decoupled from intelligence and communicability. This makes it possible to objectively position the operation itself—which induces irreversible structural updates with the environment—as "agency," even for systems lacking higher-order thinking or means of communication with human observers.

Chapter I: Introduction and Problem Statement

1.1 The Current State and Conflation of the Concept of "Agency"

In modern cognitive science, artificial intelligence engineering, and complex systems science, the term "agency" is widely used. However, its conceptual foundation remains largely abstract, often leading to academic conflation. The central issue is that agency has long been tightly bound to internal psychological attributes—such as consciousness, intention, or emotion—based on human-centric or biocentric views. Traditional models, based on humans and higher mammals, assume that the "ability to think intentionally and make purposeful choices among alternatives" is a prerequisite for agency. However, when this definition is applied, it fails to provide valid criteria for whether adaptive behaviors in insects or single-celled organisms, or operations of autonomous AI agents, should be considered to exhibit "agency." Because such psychological attributes cannot be directly observed from the outside, relying on them plunges discussions into subjective and endless debate. Furthermore, communicability with human observers is often conflated with agency, and the system's operational nature is not separated from observer convenience.

1.2 From "Agency as an Attribute" to "Agency as a Functional Process"

To address this issue, this paper proposes a fundamental shift in perspective. Specifically, it stops viewing agency as a property owned by a specific substance or organism, and reframes it as an operational process that a system carries out with its environment. Putting aside questions of physical substances or internal conscious experiences, we define agency

at the level of functional operation—how a system interacts with its environment and what state changes it produces.

1.3 Objective of the Paper and Systemic Positioning

The objective of this paper is to establish a purely systemic axiomatic framework of "Functional Agency" that is independent of hardware implementation or internal intelligence. This paper formulates the minimal operational requirements for a single-cycle operational event, forming a complementary theoretical framework with the parallel paper, *Theory of Coupled Entity Life*, which addresses the coupling dynamics of Code and Execution Substrates.

Chapter II: Three Axioms Defining Functional Agency

We introduce the following three axioms as the necessary and sufficient criteria for determining that a system demonstrates "Functional Agency." These are universal rules that prescribe the structural requirements of systemic operations.

2.1 Axiom 1: Boundary Maintenance

Axiom 1: A system forms a boundary with the environment and maintains within its boundary the structural and informational patterns that determine the system.

The Establishment of Boundaries as a Prerequisite for Operation

Forming a boundary with the environment that separates the "inside" from the "outside," and maintaining within it the structural and informational patterns that determine the system, constitutes an absolute requirement for system operation. Note that this boundary need not be a physical entity; it may also be a logical or functional boundary identifiable from the environment.

2.2 Axiom 2: Internally Initiated Discriminative Connection

Axiom 2: The system, anchored in its internal structural and informational patterns, discriminates and establishes operational contact (connection) with the structural and informational patterns of the outside (environment).

Distinguishing Passive Operation

Here, the concept of "internally initiated" stands in contrast to direct state determination from the outside (passive operation). Physical or electronic circuits whose states are unconditionally overwritten by external signals do not satisfy this axiom, as their states are directly determined by the environment. In operations exhibiting functional agency, external stimuli or data do not directly overwrite the state; rather, with internal structural and informational patterns acting as a scaffold, such stimuli or data serve as objects to be discriminated and connected.

2.3 Axiom 3: Irreversible State Change

Axiom 3: Based on the results of discrimination and connection, the system irreversibly updates its state according to its internal rules, thereby inducing changes dependent on its operational history.

Distinguishing from Reversible Motion and Physical Change (The Essence of History Dependence)

The movement of a pendulum or an elastic body that recovers from temporary deformation remains a reversible change and does not satisfy this axiom. Similarly, physical changes such as thermodynamic degradation over time (entropy increase) of the components themselves are distinct from the functional changes defined by this axiom. Irreversibility in this axiom refers to "operational history dependence," wherein the system updates its state according to its internal rules via Axiom 2 (Discriminative Connection), and the past operational history (accumulation of internal states) exerts an irreversible influence on subsequent operational rules and scaffolds (internal patterns).

Note: Initialization (reset) of internal states represents the start of a new operational cycle, which does not invalidate the history-dependent irreversibility within the operational process itself.

Chapter III: Clear Decoupling of Concepts

In this axiomatic framework, "Functional Agency" is clearly decoupled from intelligence and communicability.

Concept	Functional & Operational Nature	Relationship with the Axiomatic Framework
Functional Agency	The "operational process" itself that satisfies the three axioms	Necessary and sufficient condition
Intelligence	Computational capacity, optimization, predictive power, learning capability, etc.	Irrelevant (Optional feature)
Communicability	Whether the system can be understood or communicated with by observers	Irrelevant (Optional feature)

3.1 Decoupling "Functional Agency" from Intelligence

A major conflation in traditional theories of agency is the assumption that agency cannot exist without intelligence (such as computational or decision-making abilities) or consciousness. In this paper, we treat these two as conceptually independent. Intelligence is merely one type of complexity in the internal rules (the variety of algorithms) that leads from boundary maintenance and connection (Axioms 1 & 2) to state changes (Axiom 3).

Therefore, even if a system lacks higher-order thinking or consciousness, or has extremely limited computational capability, it is still classified as possessing "Functional Agency" as long as its operation satisfies the three axioms. Denying functional agency based on the absence of intelligence is a category mistake.

3.2 Decoupling "Functional Agency" from Communicability

Another common conflation is the view that regards the ability to converse or communicate with observers as a requirement for agency. However, even if a system's intentions are incomprehensible to observers, or the system is incommunicable (Communicability = 0), it is still classified as possessing "Functional Agency" as long as its operation satisfies the three axioms. Denying functional agency based on the presence or absence of "comprehensibility" or "communicability" to human observers is a category mistake.

Chapter IV: Application and Validation (Case Studies)

To validate the effectiveness of this axiomatic framework, we analyze operations in the following two systems against the axioms.

4.1 Biological Systems (Single-Celled Organisms)

We examine the avoidance and adaptive behaviors of *Paramecium*.

1. **Application of Axiom 1 (Boundary Maintenance)**
It forms a boundary via its cell membrane, maintaining identity without dispersing its genetic and structural patterns and its internal metabolic/reactive systems into the environment.
2. **Application of Axiom 2 (Internally Initiated Discriminative Connection)**
It uses receptors on the cell membrane surface as operational points of contact (connection), discriminating surrounding chemical substances, obstacles, or thermal gradients using receptor selectivity (structural and informational patterns) as a scaffold.
3. **Application of Axiom 3 (Irreversible State Change)**
Based on the results of discrimination and connection, it updates internal states such as membrane potential and switches the beating direction of cilia as irreversible changes under biochemical internal rules.

This analysis demonstrates that even single-celled organisms lacking higher-order thinking or human-like intelligence can be objectively determined to demonstrate "Functional Agency."

4.2 Artificial Autonomous Systems (AI Agents)

We examine an autonomous AI agent program that navigates a network to gather data or execute transactions under specified conditions.

1. **Application of Axiom 1 (Boundary Maintenance)**
Without dissolving into network data, it forms a boundary defined by its source code and execution process space, maintaining its identity.
2. **Application of Axiom 2 (Internally Initiated Discriminative Connection)**
It uses external network data or APIs as operational points of contact (connection),

discriminating retrieved information using its execution process (structural and informational patterns such as algorithms and data structures) as a scaffold.

3. **Application of Axiom 3 (Irreversible State Change)**

Based on the results of discrimination and connection, it updates internal states (such as databases or program execution status) and outputs external communications (such as sending notifications or API requests) as irreversible changes according to its algorithms (internal rules).

This demonstrates that even a software process without a physical body can be determined to demonstrate "Functional Agency" as long as it satisfies the three axioms.

Chapter V: Conclusion and Future Outlook (Parallel Contrast with the Theory of Coupled Entity Life)

5.1 Conclusion of the Paper

This paper has clarified the concept of "agency" by separating it from psychological and biocentric preconceptions, establishing an objective, systemic definition of "Functional Agency" based on the three axioms of Boundary Maintenance, Internally Initiated Discriminative Connection, and Irreversible State Change. This frames agency not as an attribute of physical matter, but as an "operational pattern." It provides a theoretical scaffold to objectively evaluate and determine autonomous behaviors in any system, regardless of the presence of intelligence or the implementation details of the hardware.

5.2 Complementary Perspectives of "Single-Cycle Operation" and "Dynamic Coupling (Life)"

This paper formulates the necessary and sufficient conditions for a "single-cycle operational event (micro-level operational unit)" in systems theory and provides criteria for evaluating single-cycle operations. In contrast, the parallel paper, *Theory of Coupled Entity Life*, describes the "overall system state (macro-level dynamics of life)" that unfolds as these three axioms are continuously manifested over time and space through the functional coupling of "Code" and "Execution Substrate." These two papers stand in a complementary relationship, constructing an integrated framework that unifies micro-level operational processes with macro-level dynamics of life.

References

1. Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press.
2. Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.
3. Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel Publishing Company.
4. Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.
5. Nakayashiki, H. (2016). *Uirusu wa ikite iru* [Viruses are alive]. Kodansha.