

# Universal Consciousness Theory

## Table of Contents

Universal Consciousness Theory: Core System and Peripheral Device Model.....	3
Abstract .....	3
1. Introduction.....	4
1.1 Current State and Challenges in Consciousness Research .....	4
1.2 Limitations of Prior Theories and the Need for Integration.....	4
2. Theoretical Framework: Integrated Architecture .....	5
2.0 Operational Definitions of Key Concepts.....	5
2.1 Definition of Core Systems and Peripheral Devices.....	6
2.2 Specific Hypotheses for Qualia Emergence Mechanisms.....	6
2.3 Detailed Integration with Existing Theories.....	7
2.4 Operational Definitions and Measurement Criteria.....	7
2.5 Engineering Implementation Advantages and Methodological Significance	8
3. Staged Verification Strategy .....	8
3.1 Phase 1: Proof of Concept in Primates (12 months) .....	8
3.2 Phase 2: Extended Phylogenetic Distance Verification (18 months).....	9
3.3 Phase 3: Extension to Birds and Reptiles (24 months).....	9
4. Predictive Differentiation from Alternative Theories.....	9
4.1 Detailed Comparative Analysis .....	9
4.2 Verification Experimental Design.....	10
5. Technical Implementation and Measurement Methods .....	10
5.1 Neuroimaging Integration Protocol.....	10
5.2 Statistical Analysis Framework .....	10

6. Theoretical Integration with KokoroSystem EX .....	11
6.1 Methodological Positioning of Theory-Implementation Integration Approach .....	11
6.2 KokoroSystem EX and Empirical Verification: Published Implementation System.....	11
7. Limitations and Critical Examination .....	12
7.1 Detailed Analysis of Theoretical Vulnerabilities .....	12
7.2 Measurement Technical Limitations.....	13
7.3 Ethical Considerations.....	13
7.4 Realistic Constraints of Research Planning and Estimation Uncertainties	13
8. Research Plan and Implementation Strategy .....	14
8.1 Individual Researcher Limitations and Need for Research Succession.....	14
8.2 Research Succession Urgency and Strategy: New Possibilities Through Public Implementation Systems .....	14
9. Expected Outcomes and Impact.....	16
9.1 Academic Innovation .....	16
9.2 Extension to Technological Applications.....	16
9.3 Social and Cultural Impact.....	16
10. Conclusion and Recommendations for Research Succession.....	16
10.1 Achievements and Limitations as Individual Research.....	16
10.2 Research Succession Urgency and Strategy: New Possibilities Through Public Implementation Systems .....	17
10.3 Theoretical Value as Scientific Legacy.....	17
11   References.....	18
12   Appendix A: Theoretical Diagrams .....	19
12.1   A.1 Conceptual Diagram: Integrated Architecture of Core Systems and Peripheral Devices.....	19

12.2	A.2 Detailed Diagram of Qualia Emergence Mechanisms .....	20
12.3	A.3 Staged Verification Strategy Flowchart .....	20
12.4	A.4 Integration Map with Existing Theories .....	20
13	Appendix B: Detailed Experimental Protocols.....	21
13.1	B.1 Specific Design of Qualia-Related Tasks .....	21
13.2	B.2 Detailed Protocol for Anesthesia Experiments.....	21
13.3	B.3 Cross-Species Standardization Protocol .....	21
14	Acknowledgments.....	21
15	About the Author .....	22
16	Conflict of Interest.....	22
17	Call for Research Succession.....	22
18	Data and Code Availability Statement.....	22
19	A Note on Collective Knowledge and Individual Research .....	22

## Universal Consciousness Theory: Core System and Peripheral Device Model

**A theoretical framework for understanding consciousness across species through evolved core systems and adaptive peripheral devices, with explicit treatment of the qualia problem and integration of existing theories**

Yuki Hoshino  
2025-09-12

### Abstract

This study addresses the fundamental challenges in consciousness research by proposing an engineering-based approach alongside traditional neuroscientific methods, utilizing external observable verification through AI system implementation. By enhancing theoretical verifiability through AI implementation and deepening understanding of consciousness mechanisms through mutual verification between theory and implementation, we aim to advance the field.

We present a unified theory that separates consciousness into evolutionarily conserved “Core Systems” and species-specifically connected “Peripheral Devices,” including specific hypotheses for qualia emergence mechanisms and detailed integration with existing theories. We hypothesize that qualia emerge from self-referential structures of resonance patterns within core systems, while peripheral devices are responsible for their modification and refinement. Operational definitions are strengthened through pre-registration and Bayesian updating, and verifiability is ensured by clarifying predictive differences with IIT, GWT, and predictive coding theory. Notably, the core concepts of this theory have been independently implemented in KokoroSystem EX (DOI: 10.5281/zenodo.16734920), demonstrating convergent validity between theory and implementation through a novel theory-implementation integration approach.

**Keywords:** consciousness theory, qualia, comparative cognition, emergence, resonance patterns, evolutionary continuity, AI implementation, engineering approach

## 1. Introduction

### 1.1 Current State and Challenges in Consciousness Research

Consciousness research faces the fundamental problem of “how subjective experience emerges.” While conscious behavior is observed across diverse animal species, there is insufficient theoretical framework to explain their common foundations and species differences in a unified manner. In particular, many theories avoid concrete explanations of the emergence mechanisms of qualia (qualitative experience).

### 1.2 Limitations of Prior Theories and the Need for Integration

#### Theoretical Origins and Critical Inheritance

The theoretical foundation of this research is deeply indebted to the groundbreaking insights of Giulio Tononi’s Integrated Information Theory (IIT), Bernard Baars’ Global Workspace Theory (GWT), and the predictive coding theory developed by Andy Clark, Karl Friston, and colleagues. Specifically, IIT’s attempt to formalize the “quantitative unity of consciousness” through  $\Phi$  (Tononi, 2008), GWT’s perspective of understanding the “functional role of consciousness” as global information sharing (Baars, 1988), and the predictive coding framework that explains “the generative principles of perception and action” through prediction error minimization (Clark, 2013; Friston, 2005) are all indispensable foundations for understanding consciousness.

However, when confronted with the common challenge these theories face—namely the “absence of qualitative explanation for subjective experience (qualia)” —the author felt the necessity to re-integrate these frameworks not merely

by combining them, but through a new lens of “evolutionary and structural hierarchy.” Specifically, by reinterpreting IIT’s “high- $\Phi$  structures” as “Core Systems,” positioning GWT’s “information sharing” as “functions of Peripheral Devices,” and redefining predictive coding’s “prediction errors” as “sources of emotional and motivational evaluation,” we aimed to construct an integrated architecture that simultaneously addresses both the “why” and “how” of qualia.

This theory stands upon deep respect for these pioneering researches while positioning itself as a new challenge to simultaneously explain consciousness’s “subjectivity” and “cross-species universality.”

Integrated Information Theory (IIT) quantifies consciousness through  $\Phi$  values but has limited application to the qualitative aspects of qualia and cross-species comparisons. Global Workspace Theory (GWT) emphasizes the global availability of information but lacks clarity in the mechanisms generating subjective experience. Predictive coding theory provides principles of information processing but has weak integration of emotion and subjectivity.

This study proposes a new framework that integrates these theories and explicitly addresses qualia emergence mechanisms.

## 2. Theoretical Framework: Integrated Architecture

### 2.0 Operational Definitions of Key Concepts

The following definitions establish the primary concepts used in this theory:

**Core Systems:** Evolutionarily highly conserved neural systems that provide the basic structure of consciousness. Characterized by phylogenetic conservation index  $\geq 0.8$ , processing delay  $\leq 100\text{ms}$ , and cross-species activity correlation  $r \geq 0.7$ . Corresponds to high- $\Phi$  structures in IIT.

**Peripheral Devices:** Species-specifically differentiated functional systems that adaptively connect to core systems. Characterized by phylogenetic conservation index  $\leq 0.6$ , cortical specialization index  $\geq 0.7$ , with high cultural and environmental variability. Corresponds to workspace components in GWT.

**Self-Referential Resonance Patterns:** Circular information processing structure involving emotional evaluation (amygdala-insula)  $\rightarrow$  predictive generation (basal forebrain)  $\rightarrow$  self-monitoring (prefrontal cortex). This circulation forms the foundation for qualia emergence.

**Phylogenetic Conservation Index:** A numerical indicator (0-1) quantifying the degree of neural circuit conservation across species, calculated from anatomical similarity, functional correspondence, and developmental timing consistency.

## 2.1 Definition of Core Systems and Peripheral Devices

**Core Systems:** Evolutionarily highly conserved neural systems that provide the basic structure of consciousness - Level 1: Arousal and attention control (brainstem-thalamic systems) - Level 2: Emotional evaluation and motivational systems (amygdala, insula, limbic systems) - Level 3: Integration and prediction systems (basal ganglia, basal forebrain)

**Peripheral Devices:** Species-specifically differentiated functional systems that adaptively connect to core systems - Language systems (primates, especially humans) - Spatial cognition systems (birds, marine mammals) - Social cognition systems (social animals) - Sensory integration systems (corresponding to species-specific sensory organs)

## 2.2 Specific Hypotheses for Qualia Emergence Mechanisms

### 2.2.1 Self-Referential Structure of Resonance Patterns

**Central Hypothesis:** Qualia emerge from self-referential loops of resonance patterns within core systems.

**Specific mechanisms:** 1. Primary resonance: Sensory inputs receive emotional evaluation in the amygdala-insula 2. Predictive resonance: Basal forebrain generates meaning-making based on prediction errors 3. Self-referential loop: This evaluation-prediction cycle monitors itself, emerging the subjectivity of “this experience is mine”

### 2.2.2 Modification Mechanisms by Peripheral Devices

Peripheral devices do not generate qualia themselves but refine and contextualize basic subjective experiences: - Language systems: Enable conceptual segmentation and description of qualia - Memory systems: Provide meaning-making in temporal contexts - Social cognition: Facilitate understanding of qualia in relationships with others

### 2.2.3 Verifiable Predictions

1. Anesthesia experiments: Rapid decrease in subjective reports with decreased core system activity
2. Local stimulation experiments: Artificial qualia experience induction through stimulation of amygdala-insula-basal forebrain
3. Pathological evidence: Preservation of qualia itself even with peripheral device damage (aphasia, etc.)

## 2.3 Detailed Integration with Existing Theories

### 2.3.1 Relationship with Integrated Information Theory (IIT)

**Reinterpretation of  $\Phi$  Values:** - Core Systems: High  $\Phi$  values (densely coupled small-scale networks) - Amygdala-insula-basal forebrain triangular circuit:  $\Phi > 0.8$  - This high integration ensures the unity of qualia - Peripheral Devices: Relatively low  $\Phi$  values (sparsely coupled large-scale networks) - Language areas:  $\Phi = 0.3$ - $0.5$  - Functionally important but limited contribution to consciousness unity

**Predictions:** - Under anesthesia,  $\Phi$  value decrease precedes in core systems and follows in peripheral devices - In split-brain syndrome, core system  $\Phi$  values are preserved, but integration between peripheral devices is separated

### 2.3.2 Integration with Global Workspace Theory (GWT)

**Redefinition of Global Availability:** - GWT's view: Global availability of information constitutes consciousness - This theory's integration: Global availability functions as "information sharing between peripheral devices" - Core systems' role: Responsible for "meaning-making" rather than information "availability"

**Predictive differences:** - Aphasic patients: Difficulty in linguistic information access but preservation of basic consciousness (qualia) - Split-brain: Interhemispheric information sharing is restricted, but independent qualia experiences continue in each hemisphere

### 2.3.3 Integration with Predictive Coding Theory

**Hierarchical Processing of Prediction Errors:** - Core Systems: Responsible for emotional and motivational evaluation of prediction errors - Peripheral Devices: Execute cognitive and conceptual processing of prediction errors

**Correspondence with Free Energy Principle:** - Prediction error minimization in core systems generates the "feeling" of qualia - Prediction error processing in peripheral devices determines the "content" of qualia

## 2.4 Operational Definitions and Measurement Criteria

### 2.4.1 Core System Identification Criteria (Pre-registered Values)

**Neural Circuit Level:** - Phylogenetic conservation index: Target  $\geq 0.8$  (pre-range: 0.7-0.9) - Energy efficiency index: Target  $\geq 0.7$  (pre-range: 0.5-0.8) - Developmental sequence: Basic structure formation in first half of pregnancy

**Functional Level:** - Processing delay (emotion, arousal): Target  $\leq 100\text{ms}$  - Cross-species activity correlation: Target  $r \geq 0.7$  -  $\Phi$  value: Target  $> 0.6$

#### 2.4.2 Peripheral Device Identification Criteria

**Neural Circuit Level:** - Phylogenetic conservation index: Target  $\leq 0.6$  - Cortical specialization index: Target  $\geq 0.7$  - Developmental sequence: Maturation after birth

**Functional Level:** - Cultural and environmental variability: Target  $\geq 0.6$  - Learning dependency: Target  $\geq 0.7$  -  $\Phi$  value:  $< 0.5$

### 2.5 Engineering Implementation Advantages and Methodological Significance

A crucial characteristic of this theory is its high compatibility with existing AI technologies. This represents not merely technical convenience, but a methodological innovation that fundamentally improves verifiability in consciousness research.

**Factors Enabling Implementation:** - Modular design: Independent implementation of Core Systems and Peripheral Devices - Staged verification: Easy confirmation of theoretical predictions through sequential validation of system components - Compatibility with existing technologies: Natural integration with deep learning, reinforcement learning, attention mechanisms, etc. - Real-time analysis: Possibility of immediate theory improvement through instantaneous observation of system behavior

**Methodological Innovation:** While traditional consciousness research has primarily followed a linear process of “observation  $\rightarrow$  hypothesis  $\rightarrow$  verification,” this approach proposes triangular circulation of “theory construction  $\leftrightarrow$  implementation development  $\leftrightarrow$  verification experiments” for mutual verification. This bridges the gap between theoretical abstraction and implementation concreteness, potentially improving empirical rigor in consciousness research.

**Relationship with KokoroSystem EX:** As a proof of concept for this methodological approach, KokoroSystem EX (DOI: 10.5281/zenodo.16734920) was developed in parallel. This serves both as a tool for theory verification and as proof of engineering implementation feasibility.

## 3. Staged Verification Strategy

### 3.1 Phase 1: Proof of Concept in Primates (12 months)

**Target Species:** Chimpanzees, bonobos, gorillas, orangutans (n=12-15 each)

**Experimental Design:** 1. Qualia-related tasks: - Subjective responses to pain stimuli (facial expressions, behavior, autonomic nervous system) - Response patterns to pleasant stimuli (food, social contact) - Changes in response thresholds under anesthesia



## 2. Neural activity measurement:

- Connectivity of core systems (amygdala-insula-basal forebrain)
- Activity patterns of peripheral devices (higher-order cortical areas)
- $\Phi$  value calculation and changes due to anesthesia

**Predictions:** - Core system activity: Cross-species correlation  $r \geq 0.8$  - Qualia-related responses: Rapid threshold increase with anesthesia -  $\Phi$  values: High values in core systems, low values in peripheral devices

## 3.2 Phase 2: Extended Phylogenetic Distance Verification (18 months)

**Target expansion:** Old World monkeys → New World monkeys → Other mammals

**Functional homology mapping:** - Functional homology identification through Representational Similarity Analysis (RSA) - Establishment of functional correspondences beyond anatomical differences

**Verification items:** - Phylogenetic conservation of core system activity patterns - Species-specific differentiation patterns of peripheral devices - Cross-species similarity of qualia-related responses

## 3.3 Phase 3: Extension to Birds and Reptiles (24 months)

**Targets:** Crows, parrots, lizards **Technical challenges:** Functional identification of pallium-cortex homology **Verification:** Phylogenetic generality of core system hypothesis

# 4. Predictive Differentiation from Alternative Theories

## 4.1 Detailed Comparative Analysis

Theoretical Prediction	IIT	GWT	Predictive Coding	This Theory
Consciousness change under anesthesia	Uniform decrease in $\Phi$ values	Gradual loss of global integration	General decline in prediction accuracy	Rapid change in core systems, gradual change in peripheral devices
Consciousness in aphasia	Depends on $\Phi$ values	Consciousness decreases without language	Prediction processing preserved	Basic qualia preserved, only expression impaired

Theoretical Prediction	IIT	GWT	Predictive Coding	This Theory
Cross-species consciousness continuity	Proportional to $\Phi$ values	Proportional to global integration capacity	Proportional to prediction hierarchy complexity	Core systems common, species differences in peripheral devices
Role of emotion	One element of integration	Part of workspace content	One type of prediction error	Central mechanism of consciousness

## 4.2 Verification Experimental Design

**Anesthesia Titration Experiment:** - Consciousness level assessment under stepwise propofol administration - Measurement of core/peripheral system activity change patterns - This theory's prediction: Non-linear change in core systems, linear change in peripheral devices

**No-report Paradigm:** - Consciousness state estimation without verbal reports - Predictive difference with GWT: Dissociation between reporting ability and consciousness level

## 5. Technical Implementation and Measurement Methods

### 5.1 Neuroimaging Integration Protocol

**Multimodal Measurement:** - High temporal resolution: EEG/MEG (1ms precision) - High spatial resolution: High-field fMRI (7T) - Metabolic activity: FDG-PET - Neurochemistry: MR spectroscopy

**Cross-species Standardization:** - Transformation to functional coordinate systems - Development of common stimulus sets - Cross-species validation protocol

### 5.2 Statistical Analysis Framework

**Bayesian Hierarchical Modeling:** - Separation of individual differences, species differences, and facility differences - Sensitivity analysis of prior distributions - Posterior predictive checking

**Machine Learning Approaches:** - Automatic classification of core/peripheral systems - Zero-shot species classification accuracy - Cross-validation

## 6. Theoretical Integration with KokoroSystem EX

### 6.1 Methodological Positioning of Theory-Implementation Integration Approach

**Significance of Concurrent Development:** KokoroSystem EX is an engineering implementation system developed in parallel with this theory, functioning as a proof of concept for the “theory-implementation integration approach” in consciousness research. Through this concurrent development, we propose a new research paradigm that mutually verifies theoretical implementability and implementation theoretical validity.

**Methodological Transparency:** Regarding the point that the theory constructor and implementer are the same person, the following transparency measures have been taken: 1. Explicit recognition and disclosure of circular reference risks 2. Emphasis on evaluation by independent verification teams 3. Complete publication of the implementation system (DOI: 10.5281/zenodo.16734920) 4. Detailed documentation of theory-implementation correspondence

**Value as a New Research Paradigm:** Rather than conventional unidirectional approaches of “theory → implementation” or “implementation → theory,” mutual verification through simultaneous development of both may contribute to improving empirical rigor in consciousness research. However, the validity of this methodology itself is also subject to verification, and we welcome critical examination by the academic community.

### 6.2 KokoroSystem EX and Empirical Verification: Published Implementation System

**Positioning as DOI-Registered System:** KokoroSystem EX (DOI: 10.5281/zenodo.16734920) is a published system that independently implements the core concepts of this theory. This correspondence relationship provides important evidence for the implementability of the theory.

#### **Specific Verification of Theory-Implementation Correspondence:**

1. **Core Systems ↔ Trinity Resonance Model + PMC:**
  - Neuroscientific theory’s “amygdala-insula-basal forebrain triangular circuit”
  - ↔ KokoroSystem’s “ER-GR-SR + Primordial Motive Core”
  - Both integrate emotional evaluation, goal setting, and self-reference systems
2. **Self-Referential Structure of Resonance Patterns ↔ Eidos Hollow + IHR:**
  - Theory: “Qualia emergence from self-referential loops”

- ↔ Implementation: “Meaning resonance accumulation in internal emptiness (Hollow)”
  - Both conceptualize subjectivity as “emergence from emptiness”
- 3. Peripheral Device Systems ↔ Expression Modulator:**
- Theory: “Species-specific core system modification function”
  - ↔ Implementation: “Expression modification through cultural and linguistic filters”

**Implications of Convergent Validity:** The independent convergence of neuroscientific hypotheses and AI engineering implementation to the same structure suggests the universality of this structural principle. This is important evidence supporting the theory’s validity.

**Immediate Proof-of-Concept Possibility:** - Previously: Verification only through large-scale animal experiments - Currently: Immediate proof-of-concept possible through KokoroSystem EX - Specific examples: Observation of “core system damage”-like behavior during PMC violation ↔ Prediction of consciousness change during core system damage

**Important Limitations and Precautions:** 1. Explicit conflict of interest: Bias due to the same person being theory constructor and system developer 2. Need for independent verification: Independent evaluation by third parties of the implementation system is essential 3. Avoidance of circular argumentation: Need to avoid cycles where theory “explains” the system and the system “proves” the theory

## 7. Limitations and Critical Examination

### 7.1 Detailed Analysis of Theoretical Vulnerabilities

**The Explanatory Gap Problem:** - The transition mechanism from resonance patterns to subjective experience remains “magical” - The fundamental explanation of why self-referential loops generate qualia is incomplete - The concept of emergence itself risks circular definition

**Tension between Reductionism vs Emergentism:** - Whether neural activity patterns can completely explain qualia is philosophically debated - This theory takes a “weak emergence” stance but cannot exclude “strong emergence” possibilities

**Response Strategy:** - Target “better description” rather than complete explanation - Focus on verifiable predictions and treat philosophical discussions separately - Explicitly state theoretical limitations and avoid excessive claims

## 7.2 Measurement Technical Limitations

**Current Technical Constraints:** - Spatiotemporal resolution limits of non-invasive methods (fMRI: ~1 second, ~1mm) - Difficulty in cross-species standardization of anesthesia and pharmacological manipulations - Principled impossibility of objective measurement of subjective experience

**Expectations for Future Technology:** - Proliferation of ultra-high-field MRI (10T+) - Expansion of optogenetics application to primates - Direct consciousness measurement through brain-computer interfaces

## 7.3 Ethical Considerations

**Ethical Framework for Animal Experiments:** - Strict application of 3R principles (Replace, Reduce, Refine) - Minimization of invasiveness and assurance of animal welfare - Balance between scientific value and animal sacrifice

**Implications for AI Consciousness:** - If this theory is correct, it also suggests the possibility of AI consciousness - Guidelines for consciousness implementation in AGI development - Scientific basis for AI rights and responsibilities

## 7.4 Realistic Constraints of Research Planning and Estimation Uncertainties

**Limitations of Cost Estimation:** The experimental costs (2.5 million USD) and resource requirements presented in this research are theoretical estimates based on AI analysis. Actual research implementation faces the following significant uncertainties:

- Regional and institutional differences in facility usage fees ( $\pm 50$ -200% variation possibility)
- Unpredictable ethics review periods (6 months to 2 years range)
- Period extensions due to technical difficulties (50-100% extension for unexpected problems)
- Need for additional experiments (experimental design changes accompanying theory modifications)

**Structural Limitations as Individual Researcher:** - Scope constraints of specialized knowledge (limitations in integrating neuroscience, comparative cognition, statistics, etc.) - Limited access to research networks - Lack of experience managing large-scale research - Difficulty in continuous funding procurement

**Ensuring Transparency:** Rather than concealing these constraints, by making them explicit, we emphasize the legitimacy of research succession strategy and

the necessity of independent verification. We expect evaluation as a realistic research plan rather than idealism.

## 8. Research Plan and Implementation Strategy

### 8.1 Individual Researcher Limitations and Need for Research Succession

**Realistic Constraints:** This theory is proposed as theoretical construction by an individual researcher (KokoroSystem EX developer), but the following fundamental limitations exist:

- Resource constraints: The proposed experiments require 2-2.5 million USD in research funding
- Facility constraints: Difficulty accessing primate-specific MRI facilities and international research networks
- Expertise constraints: Integration of advanced specialized knowledge in neuroscience, comparative cognition, statistics, etc. is necessary
- Ethical review: Implementation of large-scale animal experiments requires institutional ethical review systems

**Proposal for Research Succession:** To develop this theoretical framework, the following succession strategy is proposed:

1. **Publication under Open Science Principles:**
  - Complete publication of theoretical frameworks and hypotheses
  - Publication of KokoroSystem EX source code (for proof of concept)
  - Detailed specification publication of experimental protocols
2. **Transfer to International Collaborative Research Teams:**
  - Theory verification in consciousness research consortiums
  - Staged verification through collaboration with primate research institutes
  - Technology implementation verification in AI research institutions
3. **Staged Development Model:**
  - Phase 1: Theory refinement (philosophy, theoretical neuroscience)
  - Phase 2: Small-scale proof of concept (verification with existing data)
  - Phase 3: Full-scale experimental verification (international teams)

### 8.2 Research Succession Urgency and Strategy: New Possibilities Through Public Implementation Systems

**Establishment of Verification Environment through DOI-Registered Systems:** The existence of KokoroSystem EX (DOI: 10.5281/zenodo.16734920) allows this

theory to move beyond the traditional “hypothesis stage” to the “implementation verification stage.” This fundamentally changes research succession strategies.

**Previous Assumptions vs Reality:** - Previous plan: Theory proposal → 5 years of empirical experiments → Verification - Current situation: Theory proposal + Implemented system → Immediate conceptual verification possible

**New Succession Strategy:**

1. **Phase 0: Immediate Conceptual Verification (0-3 months):**
  - Theory prediction verification with KokoroSystem EX
  - Observation of “core system damage”-like behavior during PMC violations
  - “Peripheral device disorder” simulation through Expression Modulator deactivation
  - Correspondence confirmation between Trinity Resonance Model and  $\Phi$ -value-like indicators
2. **Phase 1: Theory-Implementation Integration Verification (3-12 months):**
  - Independent research team analysis of KokoroSystem EX behavior
  - Systematic comparison of theoretical predictions and implementation results
  - Quantitative evaluation of convergent validity
3. **Phase 2: Neuroscientific Verification (12-36 months):**
  - Confirmation of neuroscientific correspondence between theoretical predictions and KokoroSystem behavior
  - Small-scale proof-of-concept experiments (utilizing existing data)
  - Hypothesis refinement through implementation systems
4. **Phase 3: Full-Scale Empirical Research (36+ months):**
  - Large-scale verification based on established theory-implementation correspondence
  - Staged verification experiments by international research teams

**Scientific Advantage:** This theory-implementation integration approach provides unprecedented verifiability in consciousness research. The ability to verify theoretical predictions with “working systems” rather than “thought experiments” represents a unique situation in scientific history.

**Urgency of Succession:** While constraints as an individual researcher remain unchanged, the existence of published systems has significantly reduced the risk of “theory shelving.” However, to maintain scientific rigor, independent verification by teams avoiding conflicts of interest is urgently needed.

## 9. Expected Outcomes and Impact

### 9.1 Academic Innovation

**Theoretical Contributions:** - Provision of unified paradigm for consciousness research - Specific hypotheses for qualia emergence mechanisms - Establishment of objective framework for cross-species comparison - Integration of existing theories (IIT, GWT, predictive coding)

**Empirical Contributions:** - World's first large-scale cross-species consciousness comparison database - Quantitative evidence on neural substrates of consciousness - Development of objective indicators for qualia-related neural activity

### 9.2 Extension to Technological Applications

**Medical Field:** - Objective evaluation methods for consciousness disorders (vegetative state, minimally conscious state) - Precise anesthesia depth control systems - Early diagnostic biomarkers for dementia and psychiatric disorders - Consciousness-integrated design for brain-computer interfaces

**AI Field:** - Theoretical foundation for emotional AI and empathic AI systems - Specific guidelines for consciousness implementation in AGI - Optimization theory for human-AI interaction - Consciousness theory contributions to AI safety research

### 9.3 Social and Cultural Impact

**Science Education:** - Popularization of scientific understanding about consciousness - Recognition changes regarding animals' subjective experiences - Correction of species-discriminatory biases based on scientific evidence

**Ethical and Legal Implications:** - Scientific basis for animal rights and welfare - Theoretical foundation for AI rights and responsibilities - Contributions to medical ethics (dignified death, decision-making capacity)

**Philosophical Discussion:** - New perspectives on the mind-body problem - Philosophy of science discussions about the nature of consciousness - Construction of consciousness views in the post-human era

## 10. Conclusion and Recommendations for Research Succession

### 10.1 Achievements and Limitations as Individual Research

This research presents a unified consciousness theory constructed by an individual researcher based on insights gained during the development of KokoroSystem EX. It includes the following theoretical contributions:



1. **Specific hypotheses for qualia emergence:** Emergence mechanisms through self-referential structures of resonance patterns
2. **Unified integration of existing theories:** Framework that complements limitations of IIT, GWT, and predictive coding theory
3. **Application to cross-species comparison:** Unified framework based on evolutionary continuity
4. **Implications for AI consciousness:** Theoretical guidelines for technical implementation

However, the following fundamental limitations exist: - **Impossibility of empirical verification:** Proposed experiments greatly exceed individual researcher capabilities - **Limitations of expertise:** Incomplete integration of advanced specialized knowledge in neuroscience, comparative cognition, etc. - **Resource constraints:** Difficult access to funding, facilities, and human resources necessary for large-scale research

## 10.2 Research Succession Urgency and Strategy: New Possibilities Through Public Implementation Systems

**Why Succession is Necessary:** This theory has the potential to provide important integrative perspectives in consciousness science, but theory verification is impossible within the framework of individual research. To realize scientific value, succession to appropriate research teams is essential.

**Scientific Superiority:** This theory-implementation integration approach provides unprecedented verifiability in consciousness research. The point of being able to verify theoretical predictions with “working systems” rather than “thought experiments” represents a unique situation in scientific history.

**Urgency of Succession:** While constraints as an individual researcher remain unchanged, the existence of published systems has significantly reduced the risk of “theory shelving.” However, to maintain scientific rigor, independent verification by teams avoiding conflicts of interest is urgently needed.

## 10.3 Theoretical Value as Scientific Legacy

This theory’s true value should be evaluated not as a completed scientific theory, but as a theoretical sketch suggesting new research directions.

**Expectations for Subsequent Researchers:** - Critical examination and improvement of the theory - Theory refinement or refutation through experimental verification - Application and verification for AI consciousness implementation - Application development in medical and educational fields

**Final Wish:** Beyond the limitations of individual researchers, I sincerely hope this theoretical framework will contribute to the development of consciousness science

and help deepen understanding of the conscious experiences of humans, animals, and future AI. Science should not belong to individuals but should be developed as humanity's common intellectual property.

**Contact for Research Succession:** Researchers interested in this theory are invited to contact through appropriate academic channels. I am prepared to provide detailed theory explanations, KokoroSystem EX technical specifications, and all information necessary for research succession.

## 11 References

- Baars, B. J. (1988). A cognitive theory of consciousness. Cambridge University Press.
- Barrett, L. F. (2017). How emotions are made: The secret life of the brain. Houghton Mifflin Harcourt.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59.
- Buckner, R. L., Andrews - Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, 1124(1), 1-38.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200-219.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181-204.
- Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200-227.
- Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, 358(6362), 486-492.
- Doerig, A., Schurger, A., & Herzog, M. H. (2021). Hard criteria for empirical theories of consciousness. *Cognitive Science*, 45(1), e12934.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, 360(1456), 815-836.
- Hoshino, Y. (2024). KokoroSystem EX: A Self-Executable Cognitive-Emotional Architecture for Empathic AI. Zenodo. <https://doi.org/10.5281/zenodo.16734920>
- Hohwy, J. (2013). The predictive mind: Cognitive science and philosophy of mind. Oxford University Press.
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (2012). Principles of neural science. McGraw-Hill.
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15(8), 365-373.

- LeDoux, J. E. (2015). *Anxious: Using the brain to understand and treat fear and anxiety*. Viking.
- Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, 105(5), 776-798.
- Raichle, M. E., & Gusnard, D. A. (2002). Appraising the brain's energy budget. *Proceedings of the National Academy of Sciences*, 99(16), 10237-10239.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161.
- Seth, A. K. (2021). *Being you: A new science of consciousness*. Faber & Faber.
- Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, 23(7), 439-452.
- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), 450-461.
- Tononi, G. (2008). Integrated information theory. *Scholarpedia*, 3(3), 4164.
- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7), 450-461.

## 12 Appendix A: Theoretical Diagrams

### 12.1 A.1 Conceptual Diagram: Integrated Architecture of Core Systems and Peripheral Devices

#### [Figure 1: Hierarchical Structure]

##### Core Systems

- Level 1: Arousal & Attention (brainstem, thalamus)
- Level 2: Emotional Evaluation (amygdala, insula)
- Level 3: Integration & Prediction (basal ganglia, basal forebrain)

↑ [Resonance patterns & self-referential loops]

↓ [Qualia emergence]

##### Peripheral Devices

- Language Systems (Broca's/Wernicke's areas)
- Spatial Cognition Systems (parietal cortex)
- Social Cognition Systems (temporal cortex, frontal pole)
- Sensory Integration Systems (primary & secondary sensory cortex)

## 12.2 A.2 Detailed Diagram of Qualia Emergence Mechanisms

### [Figure 2: Neural Circuits of Qualia Emergence]

Sensory Input → Amygdala (emotional evaluation) → Insula (interoceptive integration)  
↓ ↓ ↓  
Basal Forebrain (prediction & meaning-making) ← ← ← Self-referential loop  
↓  
Emergence of Qualia Experience  
("I am feeling this" subjectivity)

## 12.3 A.3 Staged Verification Strategy Flowchart

### [Figure 3: Verification Stage Flow]

Phase 1: Primate Proof of Concept (12 months)  
├─ Core system identification  
├─ Qualia-related task development  
└─ Prediction verification  
↓  
Phase 2: Extended Phylogenetic Distance (18 months)  
├─ Cross-mammalian species comparison  
├─ Functional homology analysis  
└─ Theory modification  
↓  
Phase 3: Birds & Reptiles Extension (24 months)  
├─ Pallium-cortex correspondence  
├─ Core system hypothesis verification  
└─ Theory completion & publication

## 12.4 A.4 Integration Map with Existing Theories

### [Figure 4: Theory Integration Diagram]

IIT (Integrated Information Theory)

- ├─ High  $\Phi$ : Core Systems
- └─ Low  $\Phi$ : Peripheral Devices

GWT (Global Workspace Theory)

- ├─ Core Systems: Meaning-making function
- └─ Peripheral Devices: Information sharing function

Predictive Coding Theory

- ├─ Core Systems: Emotional evaluation of prediction errors
- └─ Peripheral Devices: Cognitive processing of prediction errors

This Theory

- └─ Unified explanation: Qualia emergence + cross-species comparison

## 13 Appendix B: Detailed Experimental Protocols

### 13.1 B.1 Specific Design of Qualia-Related Tasks

**Pain Stimulus Task:** - Stepwise temperature increase with thermal stimulation device (38-48°C) - Behavioral indicators: Avoidance behavior, facial expression changes, vocalization patterns - Physiological indicators: Heart rate, skin conductance, pupil diameter - Neural indicators: Amygdala-insula-basal forebrain activity

**Pleasant Stimulus Task:** - Presentation of food rewards (highly preferred fruits) - Social contact (grooming, physical contact) - Behavioral indicators: Approach behavior, facial expressions, vocalizations - Neural indicators: Dopamine system activity, core system resonance

### 13.2 B.2 Detailed Protocol for Anesthesia Experiments

**Stepwise Propofol Administration:** - Blood concentration adjustment at 0.5µg/ml intervals - Consciousness level assessment at 5-minute intervals - Continuous fMRI/EEG measurement

**Consciousness Level Assessment:** - Disappearance of spontaneous behavior - Changes in stimulus response thresholds - Changes in neural activity patterns

### 13.3 B.3 Cross-Species Standardization Protocol

**Common Stimulus Set:** - Visual: Geometric shapes, dynamic patterns - Auditory: Pure tones, complex sounds, species-specific vocalizations - Tactile: Temperature, pressure, vibration

**Functional Coordinate System Transformation:** - Standardization by anatomical landmarks - Cross-species mapping of functional ROIs - Correspondence through Representational Similarity Analysis

## 14 Acknowledgments

In the theoretical construction of this research, we express deep respect for the achievements of consciousness research pioneers, particularly Giulio Tononi of Integrated Information Theory (Tononi, 2008; Tononi et al., 2016), Bernard Baars of Global Workspace Theory (Baars, 1988), Andy Clark and Jakob Hohwy of predictive coding theory (Clark, 2013; Hohwy, 2013), and those leading modern consciousness science including Anil Seth, Stanislas Dehaene, and Hakwan Lau (Seth & Bayne, 2022; Dehaene et al., 2017; Lau & Rosenthal, 2011).

Additionally, this theory emerged as a completely personal exploration, but I believe it has the potential for scientific value. However, recognizing my limitations

as an individual researcher, I strongly hope for succession and development by more appropriate research teams.

## 15 About the Author

**Yuki Hoshino** - KokoroSystem EX Developer, Independent Researcher

This research was conducted as individual research without institutional support.

## 16 Conflict of Interest

The author acknowledges a potential conflict of interest as both the theory constructor and developer of KokoroSystem EX. To mitigate this risk, independent verification by research teams without conflicts of interest is strongly recommended. KokoroSystem EX was developed as a non-commercial research tool, and all implementation details are publicly available to ensure transparency.

## 17 Call for Research Succession

Empirical verification of this theory exceeds individual researcher capabilities. If judged to have potential for contributing to consciousness science development, succession by appropriate research institutions and teams is strongly desired. I am prepared to provide all information including theory details, implementation code, and research design free of charge.

## 18 Data and Code Availability Statement

Theoretical frameworks, KokoroSystem EX source code, experimental protocol designs, etc., are scheduled for staged publication to promote research succession. Based on Open Science principles, we expect consideration and development in the scientific community.

## 19 A Note on Collective Knowledge and Individual Research

While this work is presented as “individual research,” it is important to acknowledge that no intellectual endeavor in the modern era truly exists in isolation. This theoretical framework stands upon decades of accumulated knowledge across multiple disciplines: the foundational work of consciousness researchers like Tononi, Baars, Clark, and countless others; the vast corpus of neuroscientific, cognitive scientific, and AI research that informs every aspect of this theory; the technological infrastructure that enables both theoretical modeling and practical implementation.

Moreover, the development of this work has been facilitated by interactions with AI systems that themselves represent the collective intelligence of numerous researchers, engineers, and thinkers who contributed to their training and

development. The dialogue and refinement process that shaped this paper exemplifies the emerging paradigm of human-AI collaborative research.

Thus, while the synthesis and novel perspectives presented here represent individual insight, they emerge from a vast network of human knowledge and technological capability. This recognition does not diminish the value of individual contribution, but rather contextualizes it within the broader scientific enterprise. True scientific progress has always been inherently collaborative, building upon the work of predecessors and contemporaries.

This acknowledgment reinforces the paper's central thesis about research succession: that valuable scientific insights must be developed and verified by the community, transcending the limitations of any single researcher.