A Study on the Hierarchical Structure of Knowledge and the Cognition of Ignorance

Kunihiro Sugiyama
kunihiros@gmail.com

1 Introduction

Human knowledge possesses a multi-layered structure, and the ability to recognize one's own ignorance (metacognition) is crucial for learning and decision-making. This study proposes a continuous model of knowledge and ignorance based on four core components: **Reference** (formalized as the Truth Function T), **State** (S_k), **Self-Assessment** (R_k), and **Metacognition** (the recursive function M). This model integrates insights from **metacognition**, **knowledge representation models**, **research on the ambiguity of knowledge**, and **epistemology**. It aims to comprehensively capture three aspects that have not been sufficiently addressed in existing research: the **discrepancy between the subjectivity and objectivity of knowledge**, the **hierarchical structure of intrinsic knowledge**, and the **continuous gradation of knowledge**. By presenting a model that integrates these elements, this study seeks to deepen our understanding of the structure of knowledge and the cognitive mechanisms of ignorance.

2 Philosophical Foundation and Interpretive Notes

This section clarifies the philosophical motivation behind this paper and provides essential interpretive guidance to prevent misunderstanding of the proposed model.

2.1 Theoretical Rationale

This study is grounded in the logical structure of recursive ignorance, exemplified by the proposition **"I don't know what I don't know."** If "knowing one's ignorance" (Socratic wisdom) is a recognized concept, then logically, "not knowing one's ignorance" must also exist. And if that exists, then so must "not knowing that one doesn't know one's ignorance"—and so on, recursively.

The goal of this paper is to **mathematically formalize this recursive structure of knowledge and ignorance**, not to judge or rank cognitive states.

2.2 Descriptive Nature of the Scale

The values -1, 0, and 1 in this model function as **epistemic state descriptors**. They serve as epistemic coordinates rather than normative metrics (e.g., "good" or "bad").

Value	Meaning
1	The subject holds correct knowledge.
0	The subject lacks knowledge (ignorance).
-1	The subject holds incorrect knowledge (misconception).

Table 1: Epistemic State Descriptors

A subject in state -1 (misconception) is not normatively inferior to a subject in state 0 (ignorance); they occupy **distinct epistemic loci**. Whether one state is "preferable" to another depends on context, goals, and values—domains outside the scope of this model.

2.3 Separation of Confidence and State

A fundamental distinction in this framework is that **the state variable S_k does not encode confidence**. Confidence is handled by a separate variable R_k (Representation/Self-Assessment).

- S_k: What the subject *actually* knows (epistemic state).
- R_k: What the subject *believes* they know (confidence).

This separation is essential for capturing phenomena like the Dunning-Kruger effect, where $S_0 = 0$ (ignorance) but $R_0 = 1$ (high confidence).

3 The Reference Point: Truth Function $T(x)$

To analyze the discrepancy between what a subject believes and what is considered factual, we introduce a **Truth Function** $T(x)$.

3.1 Definition

- **Symbol:** $T(x)$
- **Definition:** A reference function that maps a proposition x to a continuous value representing its "factual status" as understood within the cognitive context.
- **Domain/Codomain:** $T : \mathcal{X} \to [-1, 1]$
  - 1.0: The proposition is considered fully accurate within the context.

- 0.0: The proposition is undefined or undeterminable within the context.
- -1.0: The proposition is considered completely contrary to the understood facts.

3.2 Ontological Neutrality

This paper deliberately adopts an **ontologically neutral** position regarding $T(x)$. We do not take a stance on whether $T(x)$ represents:

- **Objective reality** (realism): Facts exist independently of observers.
- **Phenomenal facts** (phenomenalism/relativism): Facts are constituted through the subject's cognitive context.

What matters for this methodology is that $T(x)$ serves as a **reference point** against which the subject's state $S_0(x)$ can be compared. The ontological status of $T(x)$ is a separate philosophical question beyond the scope of this paper.

This design allows users of this framework to adopt their preferred philosophical interpretation while maintaining the mathematical rigor of the model.

3.3 Connection with Epistemological Tradition

This study acknowledges the discussion of **the separation between the subjective and the objective** in epistemology. In his *Critique of Pure Reason*, Kant distinguished between phenomena (subjective experience) and the thing-in-itself (objective reality), and argued that humans cannot directly know the thing-in-itself (Kant, 1781).

Rather than claiming access to the "thing-in-itself," our model uses $T(x)$ as a practical reference point that can be operationalized in specific contexts (e.g., as expert consensus, empirical measurement, or community agreement). This distinction is an essential premise for considering the **discrepancy between the subjectivity and objectivity of knowledge**.

4 The Recursive Metacognitive Model

The cognitive structure of knowledge is modeled using a **recursive metacognitive framework**. This model is based on research findings on **metacognition** and aims to express the **hierarchical structure of intrinsic knowledge** while maintaining mathematical rigor.

4.1 Model Overview

The recursive flow of the metacognitive model can be described as follows:

1. **Layer 0 (Base Layer):**

- Input: State S_0 (Knowledge) and Self-Assessment R_0 (Confidence).
- Process: The Metacognitive Function M evaluates the alignment between S_0 and R_0.
- Output: Cognition S_1 (Self-Awareness).

2. **Layer 1 (Meta-Layer):**
   - Input: Cognition S_1 (from Layer 0) and Self-Assessment R_1 (Meta-Confidence).
   - Process: The same function M evaluates the alignment between S_1 and R_1.
   - Output: Understanding S_2 (Meta-Self-Awareness).
3. **Recursive Step:** This process continues recursively: $S_{k+1} = M(S_k, R_k)$.

Key Insight: The same function M is applied recursively at each layer, enabling arbitrary depths of self-reflection while maintaining mathematical consistency.

4.2 Core Components

4.2.1 State Variable: S_k

The state at layer k represents the epistemic condition at that level.

- **Symbol:** $S_k(x) \in [-1, 1]$
- **Meaning:**
  - 1: Holds correct knowledge (or accurate self-awareness at higher layers).
  - 0: Lacks knowledge (ignorance, or lack of self-awareness).
  - -1: Holds incorrect knowledge (misconception).
- **Important:** This is a **state descriptor**, not an evaluative score. -1 does not mean "bad."

4.2.2 Representation Variable: R_k

The representation at layer k captures the subject's **confidence** or **self-assessment** regarding their state.

- **Symbol:** $R_k \in [0, 1]$
- **Meaning:**
  - 0: No confidence ("I don't know").
  - 1: Full confidence ("I know").

4.2.3 Metacognitive Function: M

The metacognitive function evaluates the **alignment** between a state and its subjective representation.

- **Formula:** $S_{k+1} = M(S_k, R_k) = 1 - |S_k - R_k|$
- **Domain:** $M : ([-1, 1] \times [0, 1]) \to [0, 1]$
- **Output Interpretation:**
  - 1: Perfect alignment (the subject accurately recognizes their state).
  - 0: Complete misalignment (the subject is unaware of their state).

4.3 Layer Applications

4.3.1 Layer 0: State (S_0)

- **Definition:** The intrinsic state of knowledge regarding object x.
- **Expression:** $S_0(x) \in [-1, 1]$
- **Examples:**
  - $S_0 = 1$: The subject knows that "water boils at 100°C."
  - $S_0 = 0$: The subject does not know the boiling point of water.
  - $S_0 = -1$: The subject believes "water boils at 50°C."

4.3.2 Layer 1: Cognition (S_1)

- **Definition:** How accurately the subject recognizes their own Layer 0 state.
- **Expression:** $S_1 = M(S_0, R_0) = 1 - |S_0 - R_0|$
- **Meaning:** This layer captures **"knowing one's ignorance"** (Socratic wisdom) vs. **"unknowing ignorance"** (Dunning-Kruger effect).

4.3.3 Layer 2: Understanding (S_2)

- **Definition:** How accurately the subject recognizes their own Layer 1 state.
- **Expression:** $S_2 = M(S_1, R_1) = 1 - |S_1 - R_1|$
- **Meaning:** This layer captures whether the subject is aware of their own metacognitive tendencies.

S_0 **(State)**	R_0 **(Confidence)**	S_1 **(Cognition)**	**Classification**
1 (Know)	1 (Confident)	**1**	**Knowing Knowledge** — Accurate self-awareness
0 (Unknown)	0 (Unsure)	**1**	**Knowing Ignorance** — Socratic wisdom
0 (Unknown)	1 (Confident)	**0**	**Unknowing Ignorance** — Dunning-Kruger effect
1 (Know)	0 (Unsure)	**0**	**Unknowing Knowledge** — Underconfidence

Table 2: The Four Quadrants of Metacognition

4.4 The Four Quadrants of Metacognition

The relationship between S_0 (actual state) and R_0 (self-assessment) produces four archetypal cognitive patterns:

Note: The value $S_1 = 0$ for "Unknowing Ignorance" does **not** mean it is "neutral" or "acceptable." It simply means the subject **lacks awareness** of their own state. Whether this is "good" or "bad" is a value judgment outside the scope of this model.

4.5 Connection with Metacognition Research

Flavell (1979) defined metacognition as "the ability to monitor and control one's own cognitive activities." The recursive structure ($S_0 \to S_1 \to S_2$) in this study formalizes this concept mathematically.

Dunning and Kruger (1999) demonstrated the tendency of individuals with low competence to overestimate their abilities (the Dunning-Kruger effect). In our model, this corresponds to the case where $S_0 = 0$ (ignorance) but $R_0 = 1$ (high confidence), resulting in $S_1 = 0$ (lack of self-awareness).

While many existing metacognition studies focus on verifying the role of metacognition in self-learning and problem-solving processes, this study is novel in that it provides a **recursive mathematical formalization** that can extend to arbitrary depths of self-reflection.

5 The Gradation Model of Knowledge and Ignorance

In this study, knowledge and ignorance are modeled as a **continuous value** rather than as a dichotomy. This approach is based on research findings on the **ambiguity of knowledge** and aims to express the **continuous gradation of knowledge**.

5.1 Definition of the Model

- **Range**: $[-1, 1]$
  - 1: Complete correct knowledge.
  - 0: Ignorance.
  - -1: Complete misunderstanding.

5.2 Connection with Related Research

The perspective that knowledge is not always clear and complete but includes ambiguity and uncertainty is crucial for modeling knowledge as a continuous value. Zadeh's (1965) fuzzy theory provides a framework for mathematically handling ambiguous concepts, and Pearl's (1988) probabilistic reasoning provides a model for handling uncertain knowledge. These studies suggest that the continuous value model in this study is a valid approach for capturing the ambiguity of knowledge. While existing knowledge representation studies attempt to represent knowledge with clear symbols and combinations of concepts, this study is novel in that it numerically models the ambiguity and uncertainty of knowledge.

6 Analyzing Discrepancies

When subjective knowledge does not align with the reference point (factual status), this discrepancy can lead to ignorance or misunderstanding. In this study, discrepancies are analyzed at multiple levels:

6.1 Discrepancy between Reference and State (D_{TS})

- **Definition:** The discrepancy between the reference point $T(x)$ and the subject's state $S_0(x)$.
- **Equation:** $D_{TS} = |T(x) - S_0(x)|$
- **Interpretation:** Measures how far the subject's belief deviates from the factual reference.

6.2 Discrepancy between State and Self-Assessment (D_{SR})

- **Definition:** The discrepancy between the subject's actual state S_k and their self-assessment R_k.
- **Equation:** $D_{SR}^{(k)} = |S_k - R_k|$
- **Interpretation:** This discrepancy is central to the metacognitive function M. When $D_{SR}^{(0)} = 0$, we have $S_1 = 1$ (perfect self-awareness). When $D_{SR}^{(0)} = 1$, we have $S_1 = 0$ (complete lack of self-awareness).

6.3 Relationship to the Metacognitive Function

The metacognitive function M directly incorporates the discrepancy:

$$S_{k+1} = M(S_k, R_k) = 1 - D_{SR}^{(k)} = 1 - |S_k - R_k|$$

This formulation elegantly captures the insight that **accurate self-awareness** (high S_{k+1}) requires **minimal discrepancy** between one's actual state and one's self-assessment.

7 Proposed Experimental Design

To demonstrate the falsifiability and measurability of this model, we propose the **Metacognitive Alignment Test (MAT)**.

7.1 Protocol

1. **Measuring State (S_0):** Subject answers factual questions. Accuracy is measured relative to $T(x)$.
2. **Measuring Representation (R_0):** Subject rates their confidence ("Do you know this for a fact?").
3. **Deriving Cognition (S_1):** $S_1 = 1 - |S_0 - R_0|$ is calculated.
4. **Measuring Understanding (S_2):** Subject rates confidence in their Step 2 self-assessment. $S_2 = 1 - |S_1 - R_1|$ is calculated.

7.2 Validation Hypothesis

Subjects with high S_1 scores (accurate self-appraisal) are expected to perform better in subsequent decision-making tasks, regardless of their raw S_0 score. This would validate the model's claim that **metacognitive accuracy** (knowing what you know and don't know) is a distinct and measurable cognitive capacity.

8 Related Work

8.1 Metacognitive Sensitivity: meta-d'

Maniscalco and Lau developed the *meta-d'* framework for measuring metacognitive sensitivity—the ability to discriminate between correct and incorrect responses. While meta-d' focuses on **discrimination ability**, our model focuses on the **structural hierarchy** of "knowing one's ignorance" vs. "unknowing ignorance."

A key difference: existing metrics often treat "I don't know" as a failure or low confidence. In contrast, our model recognizes that $S_1 = 1$ when $S_0 = 0$ and $R_0 = 0$—meaning that **accurately knowing one's ignorance is a high metacognitive achievement** (Socratic wisdom).

8.2 Belief Functions and Uncertainty

Dempster-Shafer theory and other belief function frameworks handle uncertainty and conflicting evidence. While these approaches are valuable for modeling **epistemic uncertainty**, our model specifically targets the **metacognitive discrepancy** ($|S_k - R_k|$) that produces phenomena like the Dunning-Kruger effect.

The Truth Function $T(x)$ in our model can be interpreted within various frameworks (fuzzy logic, probability, belief functions) depending on the application context.

9 Conclusion and Future Challenges

This study constructed a recursive metacognitive model based on the hierarchical structure of knowledge. Using the components $T(x)$ (reference point), S_k (state), R_k (representation), and M (metacognitive function), we provide a mathematical framework that captures:

1. The **discrepancy between subjective belief and factual reference** (D_{TS}).
2. The **hierarchical structure of self-awareness** ($S_0 \to S_1 \to S_2 \to \ldots$).
3. The **continuous gradation of knowledge and metacognition**.

9.1 Main Results

1. We proposed a **recursive metacognitive model** that avoids the mathematical "type errors" of nested $K(K(x))$ notation while preserving the philosophical insight of recursive self-reflection.
2. We introduced the **Truth Function** $T(x)$ with **ontological neutrality**, allowing the framework to be used by researchers with different philosophical commitments.
3. We provided the **Four Quadrants of Metacognition** table, clearly distinguishing "Knowing Ignorance" (Socratic wisdom) from "Unknowing Ignorance" (Dunning-Kruger effect).
4. We proposed the **Metacognitive Alignment Test (MAT)** as an experimental protocol to validate the model.

9.2 Future Challenges

1. Empirical verification of the proposed model through the MAT protocol.
2. Simulation studies using LLM agents to validate the model's utility for AI safety (detecting "hallucination" as Unknowing Ignorance).

3. Evaluation of the impact of social and cultural factors on metacognitive patterns.

4. Application of the framework to educational interventions and decision-making support systems.

References

[1] Kant, I. (1781). *Critique of Pure Reason*.

[2] Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American psychologist*, *34*(10), 906.

[3] Dunning, D., & Kruger, J. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.

[4] Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338-353.

[5] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

[6] Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and cognition*, *21*(1), 422-430.

[7] Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

[8] Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91.