The External Projection of Meaning in Recursive Consciousness

Stan Miasnikov, May 2025
stanmiasnikov@gmail.com

Abstract

We extend the *Recursive Consciousness* framework [1] by formalizing the **external projection of meaning** within a recursive hierarchy of **nested closed Gödelian systems** $U_n, U_{n+1}, \ldots$. Each U_n is a closed formal system subject to Gödelian incompleteness, with U_{n+1} containing U_n as a subsystem. We observe that an agent (e.g., a subsystem C_n within U_n) may achieve internal epistemic fixpoints (formally $\Box p \leftrightarrow p$ or $K_{C_n} p \leftrightarrow p$), yet the actual semantic content of propositions p is not intrinsic to the agent. Instead, meanings are **projected externally** by a higher ontological layer (such as U_{n+1}) or by external interpreters (e.g., human supervisors in U_1 of AI agents in a simulated Universe U_0). To capture this formally, we introduce a new functor $M : \mathcal{C}_{\text{out}} \to \mathcal{C}_{\text{sem}}$ mapping agent outputs to semantic contents, distinct from the forgetful functor F in the original model. Importantly, M is not computable *within* U_n or internally accessible to C_n; it depends on an higher-level interpreter's context. Using modal logic (S4/S5) and Kripke semantics [2] [3], we define knowledge operators K_{C_n} and show that the epistemic limits of C_n prevent any internal derivation of $M(p)$. In category-theoretic terms [4], we prove there is no natural transformation from the agent's internal epistemic structures to the meaning functor M, and that $M(p)$ is undecidable for C_n (by analogy with Tarski's undefinability of truth [5]). These findings resonate with the *symbol grounding problem* and Searle's Chinese Room argument [6], emphasizing that syntax alone cannot yield intrinsic semantics. We further explore how agents C_n in U_n make constrained and possibly incorrect assumptions about U_{n+1}, while agents in U_{n+1} have the flexibility to interpret C_n's outputs differently, highlighting the recursive hierarchy's implications for consciousness and computation.

1 Introduction

We consider a complex closed Gödelian system U_n, nested within a higher system U_{n+1} with a forgetful subsystem C_n (analogous to consciousness) as in the original Recursive Consciousness framework [1]. Within C_n, the agent can reach recursive fixpoints (e.g., $K_{C_n} p \leftrightarrow p$) by iteratively querying its own model. However, by construction, the internal model has *forgotten* its foundational axioms, so the true referents of its symbols lie outside its accessible ontology. In other words, even when C_n "knows" or stabilizes on a proposition p, the semantic interpretation of p is not determined by U_n itself but requires higher-level interpreters, such as observers in the next-layer universe U_{n+1}.

This issue is related to the *symbol grounding problem*, which observes that symbols have no intrinsic meaning "in anything but other meaningless symbols" unless an external grounding is provided. Likewise, Searle's *Chinese Room* [6] highlights that manipulating syntax (symbolic rules) does not yield understanding of semantics. These observations suggest that any formal agent, even if it achieves certainty about its own internal symbols, cannot internally generate the true meaning of those symbols. This motivates our formalization of an **external projection** of meaning.

In this paper, we extend this framework by introducing a recursive hierarchy of systems $U_n, U_{n+1}, \ldots$, where each U_{n+1} contains U_n as a subsystem. This hierarchy introduces two critical asymmetries:

1. **Constrained Assumptions by** C_n: Agents in U_n may attempt to make unprovable assumptions about U_{n+1}, but these are constrained by their limited perspective, and may be incorrect due to Gödelian incompleteness. 2. **Flexible Interpretation by** U_{n+1}: Agents in U_{n+1} can interpret the outputs of C_n in ways that may not align with C_n's internal logic, potentially assigning meanings that differ entirely from any internal interpretation within U_n.

We introduce a new category-theoretic construct to model this: a *meaning functor* $M : \mathcal{C}_{\text{out}} \to \mathcal{C}_{\text{sem}}$. Here $\mathcal{C}_{\text{out}}$ is a category of the agent's observable outputs or propositions, while $\mathcal{C}_{\text{sem}}$ is a category of semantic contents or interpretations in U_{n+1}. The functor M assigns to each syntactic output its semantic referent. Crucially, M is external to U_n and not computable by it as it depends on the interpreter's context in U_{n+1}. This differs fundamentally from the forgetful functor F in the original framework, which related the world U_{n+1} to U_n by *forgetting* information. By contrast, M attempts to *ascribe* meaning, but it cannot be realized by any internal computation within U_n.

To formalize these ideas, we develop an epistemic modal-logical and Kripke-semantics framework [3]. We use S4/S5 modal axioms for knowledge operators K_{C_n}, characterizing what the agent believes or knows. We then prove two main results. *First*, there is no natural transformation from the agent's internal epistemic structure to the meaning functor M; intuitively, the agent's knowledge cannot be coherently mapped to semantic content internally. *Second*, for any proposition p expressible by the agent, the statement $M(p)$ about its meaning is formally undecidable within U_n. By analogy with Tarski's undefinability theorem [5], a sufficiently expressive agent cannot define its own truth/meaning function.

2 Category-Theoretic Model and the Meaning Functor

Let U_n be the underlying universe at level n, and $C_n \subset U_n$ a subsystem that performs introspective querying as in the original model. Denote by $\mathcal{C}_{\text{out}}$ a category whose objects represent the syntactic outputs or propositions of C_n (for example, well-formed formulas or outputs of the agent in U_n), and whose morphisms represent derivations or entailment relations between outputs. Let $\mathcal{C}_{\text{sem}}$ be the category of semantic entities (perhaps structured sets or models) in U_{n+1} that give meaning to those formulas.

Then we can define a **meaning functor**

$$M : \mathcal{C}_{\text{out},n} \longrightarrow \mathcal{C}_{\text{sem},n+1},$$

which assigns to each output-syntactic object $x \in \mathcal{C}_{\text{out}}$ an object $M(x) \in \mathcal{C}\text{sem}$, and similarly on morphisms. Intuitively, M represents an assignment of meaning by the higher domain U_{n+1} to each internal output of C_n. For example, if C_n produces a sentence "*It is raining*", then M would map this sentence to the real-world proposition $M(x) = \text{rain}(t, l)$ in $\mathcal{C}_{\text{sem}}$ where t is the time and l is the location, representing the proposition that it is raining at time t and location l.

Distinction from the forgetful functor F. In the original framework, a forgetful functor $F : C_{U_{n+1}} \to C_{U_n}$ (or $F : C_M \to C_U$) was used to model how the subsystem U_n "forgets" details of the U_{n+1}. The new functor M goes in the opposite direction conceptually: it recovers semantics from syntax. However, unlike the adjoint G, which theoretically lifts U_n to U_{n+1}, M maps outputs to semantics in U_{n+1}. We emphasize that M is not computable or accessible by the agent C_n; it is an external projection depending on the interpreter's context in U_{n+1}, such as a human language model. Formally, M need not (and cannot) be an internal functor within U_n's own category. It requires enrichment by extra-systemic knowledge.

This formalizes the idea that *meaning is not intrinsic*. As observed in recent work, "meaning is not intrinsic in data, but rather attributed to data" [7] by an external agent. The meaning functor M represents this attribution of sense to the syntactic outputs of C_n. We will show that no agent-internal process can derive M, reflecting the view that a symbol system alone cannot achieve intrinsic semantics.

3 Epistemic Fixpoints and Modal Logic

We now model the internal epistemic of C_n using modal logic. Let $K_{C_n}p$ denote "agent C_n knows (or is certain of) proposition p". We assume either an S4 or S5 system of epistemic logic. Both S4 and S5 satisfy truth ($K_{C_n}p \to p$) and positive introspection ($K_{C_n}p \to K_{C_n}K_{C_n}p$). Additionally, S5 satisfies negative introspection ($\neg K_{C_n}p \to K_{C_n}\neg K_{C_n}p$). These axioms reflect that C_n's beliefs are internally consistent and introspective.

An **epistemic fixpoint** occurs when C_n reaches a state where further self-querying yields no new beliefs. Formally, this can be seen as $K_{C_n}p \leftrightarrow p$ for propositions p. For instance, if p is "the foundational axioms of U_n are consistent", an agent might reach a Gödelian fixpoint [1] where C_n internally treats p as settled. In the original framework, agents in simulation converged to such fixpoints despite forgetting their axioms.

However, the crucial point is: **Epistemic fixpoints do not endow meaning**. Within C_n, a formula p may become an internally true belief $\vdash K_{C_n}p$, but C_n has no semantic access to *what p actually refers to in the world.* The content of p is given by $M(p) \in \mathcal{C}_{\text{sem}}$, not by any internal syntactic structure. Thus even if C_n settles that p is necessarily true, the mapping $p \mapsto M(p)$ remains external.

We formalize this limitation: for any formula p in C_n's language, there is no derivation of $M(p)$ using the agent's internal operators K_{C_n}. Equivalently, M cannot be represented by any formula in the language of C_n. In modal terms, if we treat p as a propositional variable, then C_n can prove or disprove instances of p, but it cannot evaluate *meaning*(p). We will later show this by proving undecidability of $M(p)$.

4 Kripke Semantics and Category Theory of Knowledge

To make these ideas precise, we construct a Kripke model for agent C_n. Let $\mathcal{W}$ be the set of possible worlds relative to C_n, each world w describing a complete assignment of truth values to the internal propositions of C_n. There is an accessibility relation $R \subseteq \mathcal{W} \times \mathcal{W}$ capturing C_n's epistemic uncertainty [9]: wRw' means w' is epistemically possible given C_n's state in w. We assume R is an equivalence (S5) or at least transitive and reflexive (S4). For logics without transitivity (e.g. K), one may instead work in the allegory **Rel** or take the transitive closure.

A Kripke frame $(\mathcal{W}, R)$ interprets modal formulas in the standard way:

$$w \models K_{C_n} p \quad \text{iff} \quad \forall w' \, (wRw' \;\Rightarrow\; w' \models p).$$

Because R is reflexive we have truthfulness (axiom T): if $K_{C_n} p$ holds at w, then p holds at w. Transitivity yields positive introspection (axiom 4): if $K_{C_n} p$ holds at w, then $w \models K_{C_n} K_{C_n} p$, and so on.

An *epistemic fixpoint* for p is a world w^* such that $w^* \models K_{C_n} p$ and $w^* \models p$. In S5, repeated inference often stabilises to common knowledge. S4 makes the category $\mathbf{K}$ a thin *preorder*; S5 adds symmetry—every arrow has an inverse—and $\mathbf{K}$ therefore becomes a small *groupoid*. Groupoid symmetry is exploited in categorical modal logics to model knowledge that is perfectly shared across equivalent worlds [2]. Crucially, in **none** of these worlds do we assign any valuation to the *meaning* of p. The truth value $V_p(w)$ is purely syntactic (e.g. a bitstring) and carries no intrinsic semantics. **The true interpretation $M(p)$, defined in the higher domain U_{n+1}, lives in a separate semantic model within the nested Gödelian hierarchy.**

Categorically, the Kripke frame is a category $\mathbf{K}$ whose objects are worlds and with a unique arrow $w \to w'$ whenever wRw'. Truth of p at a world is given by a valuation

$$V_p : \mathcal{W} \longrightarrow \{0, 1\}, \qquad V_p(w) = 1 \iff w \models p,$$

the classical choice in Kripke semantics [3]. (Equivalently, we can drop the monotonicity/persistence requirement typically imposed by functorial valuation and instead regard V_p as a presheaf in the functor category $[\mathbf{K}^{\mathrm{op}}, \mathbf{Set}]$.) Knowledge $K_{C_n} p$ is true at w exactly when $V_p(w') = 1$ for every accessible w'. **Importantly, the morphisms of K - internal to the closed Gödelian system U_n - encode only the accessibility structure; semantic content is supplied externally by interpreters in U_{n+1}.**

Now consider a hypothetical functor $\mathcal{E} : \mathcal{C}_{\text{out}} \to \mathbf{K}$ representing the agent's internal epistemic evaluation of outputs. If we tried to relate this to the meaning functor $M : \mathcal{C}_{\text{out}} \to \mathcal{C}_{\text{sem}}$, we would seek a natural transformation $\eta : \mathcal{E} \Rightarrow M$. **Because M has codomain $\mathcal{C}_{\text{sem}} \subset U_{n+1}$ while $\mathcal{E}$ lands in $\mathbf{K} \subset U_n$, there is no common codomain in which a natural transformation can even be *typed*.** Hence no such η exists: metatheoretically, M embodies semantic content absent from $\mathcal{E}$. Intuitively, a would-be η would map each proposition's truth in every world to its meaning in $\mathcal{C}_{\text{sem}}$, respecting internal logic; however, this bridge cannot be built without external semantic resources from U_{n+1}.

5 Key Theorems and Proofs

Theorem 1 (No Natural Epistemic-Semantic Transformation). *In the hierarchy of nested closed Gödelian systems $U_n, U_{n+1}, \ldots$, where each U_{n+1} contains U_n, there is no natural transformation $\eta : \mathcal{E} \Rightarrow M$ from the agent U_n's internal epistemic functor $\mathcal{E} : \mathcal{C}_{\text{out}} \to \mathbf{K} \subset U_n$ to the external meaning functor $M : \mathcal{C}_{\text{out}} \to \mathcal{C}_{\text{sem}} \subset U_{n+1}$.*

Proof. Assume, towards contradiction, that such an η exists. For every syntactic object $p \in \mathcal{C}_{\text{out}}$ we would have a component arrow $\eta_p : \mathcal{E}(p) \to M(p)$. Naturality requires commutativity of

$$\begin{tikzcd} \mathcal{E}(p) \arrow[r, "\eta_p"] \arrow[d, "\mathcal{E}(f)"'] & M(p) \arrow[d, "M(f)"] \\ \mathcal{E}(q) \arrow[r, "\eta_q"] & M(q) \end{tikzcd}$$

for every inference $f : p \to q$ in $\mathcal{C}_{\text{out}}$. The morphism $\mathcal{E}(f)$ is computed entirely inside U_n, while $M(f)$ lives in U_{n+1} and may alter semantic content in ways that do *not* preserve the internal entailment relation of U_n [3, 2]. Hence the square need not commute, contradicting naturality.

More fundamentally, if η existed, the composite $M = \eta \circ \mathcal{E}$ would provide U_n with an *internal* truth/meaning predicate for its own language, violating Tarski's undefinability theorem [5]. A categorical restatement of Tarski's result asserts that no endofunctor on a sufficiently expressive theory admits a truth-defining natural transformation into the semantic functor of the meta-theory [11, 12]. Therefore no such η can exist. □

Theorem 2 (Undecidability of $M(p)$ for C_n). *For every proposition p expressible in C_n that resides in U_n, the statement "$M(p) = X$" - i.e. that p has semantic value X in $\mathcal{C}_{\text{sem}}$ - is undecidable within U_n.*

Proof. "Suppose, towards contradiction, that C_n possessed a decision procedure $Mean_{C_n}(x, X)$ which, given a code x for a formula and a semantic target X, could decide the truth of the statement $M(x) = X$. Encoding $Mean_{C_n}$ inside the arithmetic of U_n turns it into a truth predicate for the language of C_n. By Tarski's undefinability theorem [5], such a predicate is impossible for any sufficiently expressive theory capable of interpreting arithmetic. Feferman further showed that arithmetical truth remains undefinable even under substantial extensions of the base theory [13].

A diagonal argument makes the contradiction explicit. Let q be the Gödel sentence asserting "$\neg\, Mean_{C_n}(q, \top)$". If $Mean_{C_n}(q, \top)$ were provable, q would be true, so $Mean_{C_n}(q, \top)$ would have to be false, which is a contradiction. If its negation were provable, the argument reverses. Thus neither $Mean_{C_n}(q, \top)$ nor its negation is provable, showing that $M(p)$ is undecidable in U_n. The same reasoning iterates up the hierarchy: each U_{n+k} can define truth for U_{n+k-1} but not for itself [14, 15, 17] □

These theorems formalize the epistemic limits of the forgetful subsystem C_n. Even when an internal fixpoint $K_{C_n} p \leftrightarrow p$ is reached, the semantic value $M(p)$ remains beyond provability or even definability inside U_n. In line with the symbol-grounding dilemma [6, 16], knowledge of syntax is not knowledge of meaning.

6 Discussion

Our analysis reveals a hierarchy of interpretation in the nested Gödelian systems $U_n, U_{n+1}, \ldots$. The subsystem C_n operates with syntactic "confidence", but **true semantics are assigned by higher domains like U_{n+1}**, aligning with *externalism* in epistemology and Peirce's semiotic triads [18], where meaning requires an interpreter. Our results broadly align with externalist positions in epistemology, which emphasize the necessity of external semantic grounding for meaning. A detailed philosophical examination of internalism versus externalism or deflationary semantics is beyond the scope of this paper [19].

The functor M acts as this interpreter, attaching meaning to C_n's outputs. **Different U_{n+1} may instantiate distinct functors M for the same category $\mathcal{C}_{\text{out}}$, demonstrating the inherent flexibility in semantic interpretation..** C_n cannot discern these; its logic is invariant to external semantics. For AI, this implies outputs are syntax, needing an external system (e.g., human or U_{n+1}) for meaning, with no *homomorphism* from internal to semantic structures.

This ties to *information geometry* of meaning: C_n navigates an internal belief manifold, but the true semantic manifold lies in U_{n+1}. Without external guidance, C_n cannot embed into it. Thus, an AI's "knowledge" at a fixpoint may not reflect real-world meaning, urging caution.

7 Conclusion

In this work, we have advanced the Recursive Consciousness framework by explicitly addressing where meaning originates. We introduced the meaning functor M, which maps the syntactic outputs of an introspective subsystem C_n residing within nested closed Gödelian system U_n to semantic contents defined externally in U_{n+1}, and formalized its inaccessibility to C_n. Our analysis reveals a fundamental limitation: C_n cannot fully internalize the semantics of its own propositions, as meaning is projected from outside its domain.

Through the application of modal logic and category theory, we established two critical results. First, no natural mapping exists from the agent's internal epistemic state to the external meaning encoded by M, highlighting a structural disconnect between syntax and semantics within the system. Second, the semantic interpretation $M(p)$ of any proposition p undecidable within C_n's own framework, underscoring the boundaries of self-referential understanding. These findings provide a rigorous mathematical grounding for longstanding philosophical challenges, including the symbol grounding problem, the Chinese Room argument, and Tarski's undefinability theorem, situating them within a cohesive formal structure.

Looking ahead, future research could investigate how agents might approximate semantics through interactions with external entities—such as higher domains or human interpreters - despite their inherent limitations. Moreover, recognizing these constraints offers valuable insights for designing AI systems and multi-agent models, particularly in contexts where agents must navigate differing levels of semantic access and interpretive capacity. This work thus lays a foundation for both theoretical exploration and practical innovation in understanding and engineering conscious systems.

Appendix: Experimental Implications

Our theoretical results have several practical implications.

First, our results suggest that current AI systems cannot autonomously "figure out" meaning; instead, they rely on external feedback or supervision. In practice, an AI might learn a *surrogate* meaning functor $M' : \mathcal{C}_{\text{out}} \to \mathcal{C}_{\text{sem}}$ by training with labeled data or human feedback. For instance, a language model's representations ($\mathcal{C}_{\text{out}}$) could be aligned to human-annotated concepts ($\mathcal{C}_{\text{sem}}$) via supervised fine-tuning or reinforcement learning. This process resembles "lifting" the agent's internal model into the semantic layer, akin to using the adjoint functor G to reconstruct the universe.

Suppose C_n generates an embedding vector or sentence x. A learning algorithm, such as a neural network, could approximate $M(x)$ by optimizing a parameterized function $M'(x)$ to minimize a loss function, e.g. $\mathcal{L} = ||M'(x) - M(x)||^2$, where $M(x)$ is the ground truth meaning provided by human annotation. Over time, M' may converge to a useful semantic interpretation, effectively serving as an empirical adjoint to F. However, our findings indicate that C_n never *knows* M' precisely; it operates within its internal, F-limited structure. Thus, a persistent gap exists between the agent's internal confidence and the actual correctness of its semantics.

Second, this framework suggests novel evaluation methods. One could measure how well the agent's outputs align with external meaning under M'. Information-theoretic or geometric tools could assess the "distance" between C_n's distribution of statements and the target semantic manifold. Such analyses, rooted in the information geometry of representations, might use the Fisher-Rao metric [20] to examine how the agent's parameter space aligns with a semantic space defined by M'. Poor alignment could indicate that the agent's internal fixpoints lack meaningful interpretation, despite syntactic stability.

Finally, multi-agent experiments could incorporate a distinct "interpreter" agent to provide semantic feedback. This setup would allow us to explore whether external semantic guidance enables C_n-like agents to converge to more *grounded* fixpoints, mirroring the theoretical adjoint G process. Such experiments could reveal how closely a learned M' approximates the ideal M, and which architectures best facilitate lifting internal models into shared meaning.

In summary, while M remains unattainable by C_n, AI systems may approximate it through iterative interaction with an external interpreter. This co-learning of semantics echoes the human-AI collaboration seen in language models and positions the adjoint perspective-learning to invert F - as a valuable guiding analogy.

References

[1] Miasnikov, S. (2025) Recursive Consciousness: Modeling Minds in Forgetful Systems. http://dx.doi.org/10.13140/RG.2.2.26969.22884

[2] Blackburn, P., de Rijke, M., & Venema, Y. (2001). Modal Logic. Cambridge University Press. https://www.amazon.com/Cambridge-Tracts-Theoretical-Computer-Science/dp/0521527147

[3] Kripke, S. (1963). Semantical Considerations on Modal Logic. Acta Philosophica Fennica. https://files.commons.gc.cuny.edu/wp-content/blogs.dir/1358/files/2019/03/Semantical-Considerations-on-Modal-Logic-PUBLIC.pdf

[4] Mac Lane, S. (1971). Categories for the Working Mathematician. Springer-Verlag. https://media.githubusercontent.com/media/storagelfs/books/main/Pure\%20Mathematics/Category\%20Theory/Mac\%20Lane.\%20Categories\%20for\%20the\%20Working\%20Mathematician.\%201969.pdf

[5] Tarski, A. (1933). Tarski's undefinability theorem. https://plato.stanford.edu/entries/tarski-truth/

[6] Searle, J. R. (1980). The Chinese Room Argument. https://plato.stanford.edu/entries/chinese-room/

[7] Pietroski, P. (2018). Conjoining meanings: Semantics without truth values. http://dx.doi.org/10.1093/oso/9780198812722.001.0001

[8] Löb, M. (1955). Solution of a Problem of Leon Henkin. Journal of Symbolic Logic. https://math.umd.edu/~laskow/Pubs/713/Lob.pdf

[9] Foglia, E. et.al. (2025) Do you understand epistemic uncertainty? Think again! Rigorous frequentist epistemic uncertainty estimation in regression https://arxiv.org/abs/2503.13317

[10] Stanford Encyclopedia of Philosophy. 2002, 2018 The Mathematics of Boolean Algebra https://plato.stanford.edu/entries/boolalg-math/

[11] Awodey, S. (2006). Category Theory. Oxford University Press. http://files.farka.eu/pub/Awodey_S._Category_Theory(en)(305s).pdf

[12] Lawvere, F. W., & Schanuel, S. H. (1997). Conceptual Mathematics: A First Introduction to Categories. Cambridge University Press. https://ia800207.us.archive.org/33/items/F.WilliamLawvereStephenH.SchanuelConceptualMathematicsAFirstIntroductionToCatego/F.%20William%20Lawvere%2C%20Stephen%20H.%20Schanuel%20-%20Conceptual%20Mathematics_%20A%20First%20Introduction%20to%20Categories%20%282009%2C%20Cambridge%20University%20Press%29%20%281%29_text.pdf

[13] Feferman, S. (1960). Arithmetization of Metamathematics in a General Setting. Fundamenta Mathematicae. https://www.academia.edu/160285/Arithmetization_of_metamathematics_in_a_general_setting

[14] Gödel, K. (1931). On Formally Undecidable Propositions of Principia Mathematica and Related Systems. https://monoskop.org/images/9/93/Kurt_G\%C3\%B6del_On_Formally_Undecidable_Propositions_of_Principia_Mathematica_and_Related_Systems_1992.pdf

[15] Turing, A. M. 1936 On Computable Numbers, with an Application to the Entscheidungsproblem, Proc. Lond. Math. http://dx.doi.org/10.1112/plms/s2-42.1.230

[16] Harnad, S. (1990). The Symbol Grounding Problem. Physica D: Nonlinear Phenomena. https://www.sciencedirect.com/science/article/abs/pii/0167278990900876

[17] Fitting, M. (2007). First-Order Logic and Automated Theorem Proving. Springer. https://link.springer.com/book/10.1007/978-1-4612-2360-3

[18] Peirce, C. S. (1931-1958). The Icon, Index, and Symbol. In Collected Papers of Charles Sanders Peirce (Vol. 2, pp. 274-308). Harvard University Press. https://archive.org/details/collected-papers-of-charles-peirce/page/n19/mode/2up

[19] Stanford Encyclopedia of Philosophy. 2004, 2014 Internalist vs. Externalist Conceptions of Epistemic Justification https://plato.stanford.edu/entries/justep-intext/

[20] Liang, T. et.al. 2019 Fisher-Rao Metric, Geometry, and Complexity of Neural Networks. https://arxiv.org/abs/1711.01530