Model Theory, Topos Theory, and the Donut of Attention

The fields of model theory and topos theory emerged in the 20th century as powerful abstract frameworks for "models," "spaces," and "truth." Model theory arose in logic and mathematics as the study of formal languages and their interpretations by set-theoretic structures. Grothendieck's work in algebraic geometry led to the notion of a topos – initially a category of sheaves on a site – which Lawvere and Tierney later axiomatized as an "elementary topos" with an internal logic . These parallel developments reframed how diverse disciplines conceptualize models and spaces. In each case, truth and semantics shift from an absolute external standard to features of the model or space itself. Model theory, building on Tarski's definition of truth, treats a model as a structure making sentences true . Topos theory generalizes the notion of space: a topos is like an "alternative universe" of sets with its own internal logic (often intuitionistic rather than Boolean).

In logic and semantics, these ideas converge: every first-order language has models in set-theoretic structures (model theory) and can also be interpreted in the internal language of a topos. For example, the semantics of natural language can be seen as a model-theoretic mapping of sentences to truth-values in a structure . In categorical logic, a topos carries a "subobject classifier" (an internal truth-value object) so that logical operations (and even constructive set theory) can be carried out internally . In this way models become points or sections in generalized spaces, and truth is a morphism within a space. Both theories thus provide languages for representing structure and meaning: model theory via sets-and-relations interpretations, and topos theory via objects-and-morphisms in a category with logical structure .

Historical Evolution: From Tarski and Grothendieck to Isham

Early milestones set the stage. In the 1930s, Gödel and Tarski showed that formal theories have nonstandard models and defined truth semantically . By the 1940s–50s Mal'tsev (in Moscow), Tarski (USA), and Robinson (UK) realized that logical metatheorems could prove genuine mathematical theorems about algebraic structures. Robinson, addressing the 1950 International Congress, even noted that symbolic logic was fulfilling a Leibnizian dream of unifying algebra and geometry . In fact, by 1954 Tarski was already calling this new subject the theory of models . Hilbert's earlier work on models of geometry foreshadowed this: he constructed non-Euclidean models by interpreting axioms abstractly. Meanwhile, in algebraic geometry the absence of enough "points" in varieties motivated Grothendieck (circa 1960) to treat schemes as functors – leading him to invent the Grothendieck topos (a category of sheaves) to generalize space . Lawvere and Tierney then recognized that Grothendieck's idea could unify logic and topology: they axiomatized an elementary topos with a subobject-classifier (generalized truth values) and cartesian-closed structure .

Thus, by the late 20th century, two parallel landscapes had formed: model theory (now a deep branch of mathematical logic) and topos theory (a bridge between geometry and logic). Both were recognized as foundational. In fact, the topos framework became sufficiently general that Hilbert's intuitionistic logic (once at odds with classical logic) found a home: in each topos there is an internal Heyting-algebra of truth-values and each topos can be seen as a setting for constructive mathematics . This categorical-logic fusion influenced computer science as well: Lawvere's program (in Lambek–Scott 1986) showed that higher-order logic could be interpreted in any topos, making toposes into "alternative set theories" or universes of discourse .

More recently, Chris Isham and Jeremy Butterfield (1990s–2000s) applied topos ideas to physics. They noted that quantum theory's paradoxes (e.g. the Kochen–Specker theorem) reflect a failure of classical truth-value assignments. By working in a suitable topos of presheaves one can restore a form of "neo-realist" semantics for quantum observables . In fact, Isham, Döring, and others suggest that topos physics – where each physical theory lives in its own topos – could be the framework for quantum gravity . For example, Tore Dahlen (2011) explicitly investigated how loop quantum gravity (LQG) might be recast in a topos-theoretic language, using "Bohrification" to associate algebraic observables in LQG with sheaf structures . As one recent survey notes, "one of the main motivations behind so-called topos physics…is to provide a framework for new theories of quantum gravity" . In this sense, the historical arc runs from Hilbert's geometric models and Tarski's semantic truth definitions, through Grothendieck's and Lawvere's categorical spaces, to a topos-based quest for spacetime quantum gravity.

Models, Truth, and Semantics

At its core, model theory asks: What does it mean for a statement to be true in a structure? A model is simply an interpretation (a set or collection of sets with relations/functions) that makes sentences true. Tarski's paradigm (1933) defined "truth" in a model relative to a recursive definition on formulas . Thus model-theoretic truth is "truth in a particular interpretation" . For example, the sentence "There exists an x such that…" is true in a structure I precisely if the interpretation of the predicate and function symbols in I makes that existential sentence true. In this view, different models (even of the same theory) can give different "realities." Hilbert's 1899 demonstration that non-Euclidean geometries exist came from showing that different models of Euclid's axioms have different truths (parallel postulate holds or fails). Model theory systematized this: properties like compactness, completeness, categoricity, and various degrees of complexity (stability, simplicity) classify theories by the behavior of all their models. Thus model theory itself became a meta-mathematical study of what models a theory admits and what those models "look like."

Model theory also bridges to semantics of language. The Tarskian notion extends to natural and formal language: assigning meanings to sentences by mapping them into a model. In computational linguistics, Montague semantics famously uses model theory: a grammar is an algebra of expressions, and each expression's meaning is defined as an element or set in a semantic model (usually sets, functions on a domain) . Johnson-Laird's "mental models" in cognitive psychology similarly treat beliefs as internal models that "explain" language understanding and reasoning. In all cases, models serve as cognitive-semantic structures that ground symbols in concrete (or at least mathematical) entities. In AI and formal systems, this is echoed: a programming language has an abstract "state space" and its programs are modeled as transformations on that space. As one logic-in-AI survey notes, "logic can, for instance, provide a specification for a programming language by mapping programs to the computations that they should license, enabling proofs that these computations conform to certain standards." . In effect, the program's execution model is a structure, and formal verification checks that program actions satisfy logical specifications within that model.

Above all, model theory reframes truth as a relation between sentence and structure: "I is a model of sentence S" means "S is true in I" . This relativizes meaning: different models give different truths. Philosophically, this led to robust semantics and multiple "worlds" in modal logic. In the metasemantic and philosophical view, one studies the space of all models of a theory to understand meaning and consequence . Thus, in logic and mathematics models are primary objects of study, and truth is a property internal to a model.

Topoi and Generalized Spaces

While model theory fixes a language and varies the structure, topos theory generalizes the notion of structure itself. A Grothendieck topos is a category of sheaves on some site (think of gluing local data over a space). Lawvere–Tierney axiomatized this in 1964 as an elementary topos: a category with finite limits, power objects, and a distinguished "truth object" \Omega classifying subobjects (generalizing the set \{\text{true},\text{false}\} in Set). In practical terms, every topos behaves like a universe of variable sets: one can do virtually all of mathematics inside it, but possibly without the law of excluded middle (so logic may be intuitionistic). In such a universe, "truth" is itself an internal object (a subobject of the terminal object) and logical connectives arise from categorical operations on \Omega.

For example, the category \mathrm{Sh}(X) of sheaves on a topological space X is a topos. It "remembers" a lot about X: points, open covers, and local data all become part of the category. As Baez explains, "a topos can also be seen as a way of talking about a space," even if it is not literally a space of points . More generally, any Grothendieck topos is an elementary topos (a "universe" with its own logic) . This bridges geometry and logic: axioms of geometry become properties of objects in the topos. In fact, many logical theories (even set theory variants, type theory, and real analysis) can be interpreted in an appropriate topos, often yielding "forcing" or independence results.

From the topos perspective, a space may have no points at all (e.g. locales, which are "pointless" topological spaces). One famous slogan is "points come later": one first builds a topos of sheaves, then any notion of point of that topos is derived, not fundamental. So a topos serves as a kind of generalized space or universe. For instance, Lawvere and Tierney showed that models of set theory (including Cohen's forcing models) correspond to certain toposes where the subobject classifier \Omega encodes alternative truth values . In categorical logic, the internal language of a topos with \Omega a Heyting algebra realizes a fully general (usually constructive) set theory or type theory . Saunders Mac Lane emphasized that category theory and toposes reveal the ubiquity of these concepts: once established, "topos" unifies many mathematical worlds .

To summarize, model theory fixes a language and studies its models, while topos theory fixes a categorical structure and explores its internal logic and geometry. Both are frameworks for representing relationships: model theory uses Tarski-style interpretations, and topos theory uses sheaf-like objects and morphisms. Notably, a first-order structure can be seen as an object in some topos, and conversely each topos has an underlying "category of models" of its internal language. This duality underlies much modern semantic and categorical logic.

Physics: Quantum Gravity, Field Theory, and Holography

In physics, these abstract ideas find novel analogues. Loop Quantum Gravity (LQG) discretizes spacetime with spin networks (graphs labeled by group representations). These networks have the topology of loops and toroidal connections. Mathematically, the state space of LQG is built from holonomies on loops, a kind of gauge field on a (generalized) space of connections. Tore Dahlen shows that one can treat LQG's C*-algebra in a "Bohrified" topos sense, blending LQG with topos techniques . Thus LQG, which inherently uses combinatorial/categorical data (graph fusion), admits reinterpretation via sheaf-like structures over its space of connections. More conceptually, spin foam models of quantum spacetime resemble 2-categories or higher-categorical gluing of building blocks, hinting at topoi of histories.

Another frontier is the topos approach to quantum theory pioneered by Isham and Butterfield. Quantum observables do not all commute, so one cannot assign a single global truth function. Their insight was to work in a topos of presheaves over the commutative subalgebras of the algebra of observables. In this topos, each quantum quantity has a spectrum object, and "truth" values become sieves (generalized sets of contexts). In effect, they replace the usual Hilbert space probabilistic picture with a logic-valued picture internal to a topos. As Isham & Butterfield state, "we discuss some ways in which topos theory… can be applied to interpretative problems in quantum theory

and quantum gravity.” . For example, the Kochen–Specker theorem (no global valuation) becomes simply the statement that no global section of a certain presheaf exists; but locally (in each context) truth-values still live in the topos.

Quantum Field Theory and holography also echo these themes. The AdS/CFT correspondence shows that a gravity theory in a “bulk” space is fully encoded by a field theory on its boundary – a holographic duality. Conceptually, one can think of the bulk as a “space” and the boundary CFT as a model living on that space. The holographic principle suggests that what we think of as volume degrees of freedom really live on a lower-dimensional space. In a topos-like spirit, one may view the boundary as providing an internal “sheaf” of information that reconstructs the bulk geometry. While this is still speculative, physicists often note analogies between holography and logical dualities (e.g. bulk fields as global sections of some sheaf). Some have even applied ideas from noncommutative geometry and topos theory to understand space at the Planck scale. In all, high-energy physics exploits generalized spaces (like loop networks and holographic screens) and novel logics, suggesting that model-theoretic and topos-theoretic structures underlie spacetime and fields.

Cognition and Linguistics: Mental Models and Semantics

In cognitive science and linguistics, model-like and topos-like frameworks arise in how mind represents meaning. Mental models (Johnson-Laird et al.) are internal structures that mirror external reality, much like logical models mirror theoretical axioms. Language semantics is explicitly model-theoretic: a grammar provides syntactic rules, and semantics assigns to each sentence a condition on a model so that truth can be evaluated. As one logic encyclopedia notes, such an interpretation “explains what objects some expressions refer to and what classes some quantifiers range over” . This is exactly how Montague semantics treats natural language: words denote elements or sets, sentences denote true/false in a model.

Category theory also enters cognitive semantics. For example, Lambek and Scott showed that types and grammatical compositions can be treated categorically, and natural language grammars resemble categories (Lambek calculus). Conceptual metaphor theory (Lakoff) informally treats concepts as mappings between spaces – though not formalized, this resonates with viewing cognition as functors or sheaves gluing contexts. Some cognitive scientists invoke conceptual spaces (Gärdenfors) where concepts are points or regions in a geometric (often topological) space, akin to objects in a category of cognitive states. While strict topos models of mind are not yet mainstream, there is growing interest in sheaf-theoretic models of distributed knowledge and context.

The brain’s physiology even shows fractal and holographic patterns. Empirically, neurons are highly fractal in shape: one study of rat neurons finds that “their dendrites fork and weave through space in a fractal-like manner… [and] their [fractal] D values reflect a network cooperation that optimizes… constraints.” . In other words, neurons exploit fractal geometry to maximize connectivity with minimal cost . This suggests the brain’s wiring uses multi-scale patterns – a kind of natural “fractal topology” that may facilitate complex information flow. Holographic models of cognition have also been proposed. Karl Pribram’s holographic brain theory (1980s) imagined memory as a hologram stored in interfering waves. Recently, Nishiyama et al. (2024) develop a quantum variant: “We investigate… the holographic brain theory introduced by Pribram to describe memory in the human brain. …We adopt binary holograms to manipulate optical information… [in a] hierarchical model.” . In short, cognition seems to leverage nested and distributed representations (like fractals and holograms) rather than simple localist models. Language and meaning likewise are fluid and context-dependent, matching the flexibility of sheaf-like and model-theoretic semantics.

Systems Theory and AI: Formal Models and Computation

In systems theory and AI, model-theoretic ideas manifest in formal verification, semantics, and abstract interpretation. The semantics of programming languages is often given by mathematical models: e.g. the states of a program form a Kripke structure or operational model, and one proves properties (invariants, safety) by mapping programs into logical models . Model checking explicitly checks whether a state-transition model satisfies a temporal-logic specification. Abstract interpretation (Cousot) is essentially model theory for program analysis: it constructs abstract models of program behavior to infer properties like possible values or termination. In each case, the program’s behavior is represented within a mathematical “model” and logic is used to analyze it.

Category theory underlies many modern computational models. The Curry–Howard correspondence identifies proofs with programs and logical formulas with types. Type theory (especially dependent type theory) can be seen as an internal language of a topos. This has given rise to homotopy type theory (HoTT), where types are homotopy spaces and equality is a path; here geometry and logic fully merge. HoTT exemplifies how an abstract categorical logic can become a computational framework (programming languages like Agda embody dependent types). Other categorical frameworks include game semantics (where proofs/strategies play games), sheaf semantics of concurrent systems, and topos-theoretic models of distributed computation.

this light. Knowledge graphs and ontologies are essentially directed labeled graphs (categories) encoding relations. Neural network architectures (e.g. grid-like CNNs) often have symmetries and topologies (torus convolutions for periodic boundary conditions). Interestingly, some AI researchers explicitly try to embed data in tori or hyperbolic spaces to capture hierarchical structure. While connectionist models use geometry and dynamics, symbolic AI relies on discrete models and logic. However, as the SEP on logic in AI observes, even practical systems often incorporate logical ideas: “a software application can be said to implement a logical formalization… when logical ideas informed parts of the software development process.” . In both cases, model theory and category theory provide the languages for describing state spaces, transformations, and emergent properties of computational mind-like systems.

Cross-Disciplinary Metaphors: Donuts, Fractals, Holograms

To bridge these formalisms with intuitive imagery, researchers often invoke geometric metaphors. Toroidal or “donut” shapes arise in dynamical systems and topology: a torus can represent a 2D periodic domain or a loop of feedback. In neural network models, one sometimes uses toroidal grids to avoid boundary effects. Metaphorically, one can imagine attention or consciousness as flows on nested rings – loops of processing that intersect. For example, a simple recurrent attention loop might be pictured as a toroidal circuit of information. The notion of a nested toroidal dynamics (as in the Donut of Attention project) suggests multiple layers of loops: a small torus of neural firing curves embedded in a larger torus of cognitive focus, and so on. This resonates with the idea of spirals of thought in cognitive science and with attractors in phase space that can be toroidal. Although formal models of consciousness are still rudimentary, one could envision a hierarchy of categories (or topoi) each looping into the next, a categorical model of layered awareness.

Fractality is another powerful metaphor. Fractals have self-similarity across scales, and indeed the brain and mind exhibit multi-scale structure. As noted, neuron dendrites are literally fractal branching . Cognitive “fractal” suggests that patterns of thought, language, or network connectivity repeat at different scales – from local circuits to large-scale brain networks. Some theories of perception even propose that we recognize fractal patterns more easily due to evolutionary bias. A recent review argues that fractals serve as “cognitive scaffolding rather than ontological statements”: they are tools the brain uses to impose order on complex inputs . In information theory terms, a fractal code might allow the same symbol sets to represent structure at many levels.

Holography provides a metaphor of global-local duality. In physics, a hologram encodes a 3D image on a 2D surface; in AdS/CFT the entire bulk geometry is encoded on the boundary. Pribram’s holographic brain theory likewise held that memory is stored in interference patterns (wavelets) so that each piece of memory (like a fragment of hologram) contains the whole information. The recent quantum hologram model suggests that the brain may use wave interference in water molecules (microtubules) to achieve coherent memory storage. Conceptually, a holographic mind implies that any local observation reflects global patterns. If attention is holographic, then focusing on a detail implicitly references the entire context. This dovetails with topos ideas: in a sheaf, local data (sections over patches) piece together the whole. One can imagine consciousness as stitching together local percepts into a global picture, just as a topos stitches local truth into a global logic.

Thus the metaphors of donut, fractal, and hologram all capture aspects of the same vision: information and meaning flow through nested, looping, multi-scale networks. A torus can represent a cycle of thought, fractal patterns can represent nested levels of analysis, and holographic patterns can represent integrated wholes. These are not formal models per se, but they guide how one might map high-level cognitive phenomena onto mathematical structures.

Toward a Unified Architecture for Attention

nested toroidal flows. Model theory and topos theory suggest concrete ways to realize this vision. In essence, we seek a “multi-topos” or “multiverse” of attention: each level of focus has its own internal logic (like an internal topos), yet these are linked by morphisms or interpretations (models between theories). One can imagine each layer of consciousness as a category of models (a topos) whose objects are mental states, whose morphisms are transitions of thought, and whose internal logic captures what is considered true or salient at that level. Moving from one layer to another involves “interpreting” one model in another (as in model-theoretic interpretation of one theory in another).

mental model (attentive context) certain statements are true. Topos theory reframes a space of attention as a flexible environment with its own logic. The intersection of these suggests a structural semantics of mind: every belief state is a model in some topos, and every learning or insight is a morphism between toposes. The hierarchy of attention could thus be a tower of categories (a 2-category, or higher stack), where the internal logic at each level governs that level’s “truths” (for example, lower-level sensory constraints vs higher-level semantic truths). Such an architecture naturally accommodates symbolic resonance: symbolic meanings would be objects whose interpretations resonate across multiple layers, much like a sheaf section that extends over overlapping open sets.

In more concrete terms, formal verification and abstract interpretation hint at how this could work in a computational model of mind. For example, imagine a simplified cognitive system where each “module” has a formal model of its world (a Kripke structure or topos) and communication between modules is interpretation of one model in another. Attention flow could then be captured by functors or sheaf restrictions that pass information between contexts (e.g. focusing attention corresponds to restricting a presheaf to a smaller open set). Category theory already provides tools like fibrations and indexed categories precisely for representing “systems of systems.” For instance, a fibration of topoi could represent a global structure of attention that fibers into local sub-topoi of perceptual modules.

This framework could also incorporate geometry in the form of tori and spin networks. A spin network in loop gravity is a graph labeled by representations, which as a category encodes adjacency and connectivity. Similarly, a toroidal attractor in dynamical systems is a state space homeomorphic to a torus, which can model coupled oscillations. If attention is rhythmic and cyclical (as neuroscience suggests with neural oscillations), then toroidal state spaces might naturally arise. Higher analogues (spin foams, 4D toroidal volumes) might model layers of time or creative insight. The notion of a “holographic surface” can analogize how a conscious snapshot (a high-dimensional pattern of brain activity) is encoded in lower-dimensional patterns.

Key related frameworks that could be integrated include: • Game semantics: Models meaning as games between Agent and Environment; attention could be the “game” of selecting stimuli (with each move as a focus shift). • Sheaf theory: Already hinted above as a way to

glue local beliefs into global coherence. • Homotopy Type Theory: Treats types as spaces and equivalences as paths; this could allow multiple, equivalent "paths" of thought representing the same concept, echoing human concept flexibility. • Information theory: Measures like entropy or integrated information (Tononi's Φ) could quantify the "richness" of a topos or model. Attention might optimize information flow across layers.

All of these remain speculative. But the interdisciplinary trend is clear: abstract structures from logic and geometry are being repurposed to model cognition. For example, graph neural networks are essentially category-theoretic: data live on nodes (objects) and messages (morphisms) propagate. Quantum cognition even tries to use Hilbert-space models (non-Boolean logics) for human decision making. The "Donut" metaphor invites us to see these as slices of a larger toroidal architecture: feedback loops (a torus), self-similar patterns (fractals), and holistic interdependence (holography).

Future Speculations: Consciousness, Time, and Creativity

Looking ahead, one can imagine further extensions. Multi-layered consciousness might be modeled by a tower of topoi or an (\infty,1)-category of belief states. Homotopy type theory suggests that identities between thoughts can themselves vary, offering a way to encode subjective uncertainty or creative reinterpretation. Creative time – the feeling that time stretches or contracts during novel thought – might find mathematical analogues in noncommutative geometry or fractal time models, where scale changes affect the flow of time. One could even speculate that each "moment" of attention is a topos object, and the morphism of time is a functor acting on a tower of such objects.

Philosophically, this resonates with process philosophy (Bergson's durée) and dynamic systems: consciousness is not static but a continuous transformation. Category theory can formalize continuity by considering limits, colimits, and cohomological invariants that capture global structure of changing states. Cognition might then be seen as computing fixed points or attractors in these higher spaces. Symbolic resonance – the way certain ideas recur and reinforce across contexts – could correspond to invariant subobjects or natural transformations that persist under many functors.

In summary, model theory and topos theory offer rich metaphors and tools for a unifying architecture of mind. They teach us that models, spaces, and truths are context-dependent. By thinking of attention as flows on nested tori, and cognition as sheaf-like gluing of information, we create a language where abstract mathematics and subjective experience can meet. While much remains conjectural, these interdisciplinary bridges hint at a future where logic, geometry, and metaphors like the donut coalesce into a coherent science of consciousness and creativity.

Sources: Authoritative accounts of model theory and topos theory provide the historical and conceptual basis . Applications to physics (quantum gravity, holography) and cognitive models are discussed in recent literature , and connections to computation and semantics are noted in logic/AI surveys .