TOWARD A UNIFIED MATHEMATICS OF CONCEPTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Concepts are commonly defined as abstract, compact representations of knowledge and treated as basic units of intelligent behavior. Yet, cognition, psychology, and AI lack a shared mathematical language for them. Modern systems represent concepts as vectors, distributions, symbols, graphs, and other structures, but these formalisms are typically treated as competing rather than as solutions to a common problem. We propose an operation-based view that evaluates mathematical frameworks by the conceptual operations they support, identifying thirteen operations (including similarity, composition, generalization, and grounding) that recur across cognition, psychology, and AI. We show that ten frameworks embody distinct commitments to concepts as self-contained content, relational structure, or evolving process, and that these commitments determine which operations each supports naturally. For example, vector-based models facilitate graded similarity and generalization but struggle with explicit composition, whereas symbolic models support composition but offer but generalize poorly. No single framework we examined naturally supports all operations without extension. We test this account empirically using categorization as a case study, operationalizing nine theories on the same items against human judgments. Despite addressing the same conceptual question, the theories produce different procedures and results, demonstrating that mathematical commitment shapes what a theory can explain. We call for hybrid formalisms that treat content, relation, and process as jointly primary.

1 INTRODUCTION

> "One of the pleasures of looking at the world through mathematical eyes is that you can see certain patterns that would otherwise be hidden." – *Steven Strogatz*

Concepts are how intelligence compresses the world. A single concept like *zebra* lets us recognize a novel instance despite changes in pose or lighting, generalize to related categories like *horse*, understand that *horse* + *stripes* = *zebra*, and combine it with others into new thoughts like a *zebra tea party*. Modern LLMs, too, appear to form internal representations that support analogy, generalization, and other documented concept operations. This hints that concepts are as fundamental to artificial systems as to biological ones.

While researchers agree that intelligent systems converge to similar abstractions, there is **no agreed mathematics of concepts** (LawrenceC & Jenner, 2023). Modern AI represents concepts as vectors, graphs, symbolic predicates, probabilistic variables, sparse features, and many other forms (Johnston & Fusi, 2023). These are often viewed as competing representations, each with their own strengths and limitations. We argue that this framing is premature. **Before asking which mathematics best represents concepts, we must first ask what it is we wish to represent** (Kothari et al., 2026).

Scientific progress has often been driven by finding the right mathematical language. Differential equations transformed mechanics, probability formalized uncertainty, and graph theory reshaped the study of networks. Mathematics does not merely describe phenomena, but rather determines which structures become natural to represent and which questions become possible to ask. If concepts are the fundamental units of intelligence, then the mathematics used to represent them will shape the kinds of intelligent systems we can imagine and build.

We propose that the fundamental object of study is not a particular representation of concepts, but the capabilities that concepts enable. Intelligence does not merely store concepts; concepts enable

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

intelligent systems to compare and categorize, abstract and generalize, compose and reason, adapt meaning to context, connect knowledge to the world, and acquire and revise conceptual knowledge. At the same time, mathematical frameworks differ in whether they treat concepts as entities with intrinsic structure, as elements defined through their relations, or as processes that emerge and change over time. **A mathematics of concepts must therefore support the operations concepts support**.

We derive these operations by identifying capabilities that recur independently across comparative cognition, cognitive psychology, and AI, and use them to evaluate existing mathematical approaches. We argue that different mathematical languages capture complementary aspects of conceptual intelligence and together define a new research agenda: the mathematics of concepts.

This paper makes the following contributions: (i) an operation-based framework that defines a mathematics of concepts by the conceptual operations it supports rather than by a preferred representation, identifying thirteen operations that recur across comparative cognition, cognitive psychology, and AI (§ 2); (ii) a systematic comparison of ten mathematical frameworks against these operations, showing that each framework's profile of strengths tracks a prior commitment to content, relation, or process (§ 4, 5, 6, Table 2); (iii) an empirical categorization test-case operationalizing nine of these frameworks on the same dataset against human judgments, showing how mathematical commitment shapes both the procedure and the result for a single conceptual question (§ 7, Table 5); and (iv) a diagnosis of why no existing framework spans content, relation, and process jointly, and a concrete agenda for hybrid formalisms that treat all three as jointly primary.

2 What Should a Mathematics of Concepts Explain?

Every mathematical theory develops around the assumptions and operations it is designed to capture. Thus, we ask: *what do intelligent systems actually do with concepts?* This *operation-based view* characterizes concepts by the cognitive and computational operations they support, rather than by their representations alone. This would define the requirements for any mathematics of concepts.

2.1 Fundamental Conceptual Operations

Table 1 introduces the 13 concept operations we identified to recur independently across comparative cognition, cognitive psychology, and AI. Although these traditions differ substantially in their methods, assumptions, and explanatory goals, they repeatedly converge on a common set of capabilities that concepts enable across intelligent systems.[1]

Consider the concept *zebra* : Abstraction extracts features shared across particular zebras, such as their characteristic stripe pattern, while discarding incidental detail such as which zoo or habitat they are observed in; analogy maps the relation between *horse* and *zebra* onto the relation between *dog* and *wolf*, a domesticated-to-wild correspondence; categorization assigns a new instance to the *zebra* category (e.g., a new zebra at the local zoo); communication allows the word "zebra" to reliably evoke the same concept in a listener as in the speaker; composition combines the simpler concepts *horse* and *stripes* into *zebra*, such that the meaning of the whole is predictable from its parts; contextualization changes its interpretation across uses such as an actual *zebra* versus a "zebra crossing"; discrimination distinguishes *zebra* from the closely confusable concept *horse*; generalization applies *zebra* to a previously unseen subspecies, such as the imperial zebra; grounding connects the concept to sensory experience of actual zebras encountered in the world (the way they feel, smell, etc.); inference derives that a particular zebra is a herbivore from its membership in *zebra*, without observing this directly; learning acquires the concept *zebra* in the first place from repeated exposure to instances, for example through picture books or zoo visits; similarity judges *zebra* as closer to *horse* than to *giraffe*; and revision updates *zebra*'s boundaries on learning that Okapi, despite a very similar stripe pattern, is not a type of *zebra*.

3 Evaluating Mathematical Frameworks

The multifaceted nature of concepts suggests that mathematical frameworks will exhibit systematic strengths and weaknesses across conceptual operations. We test this hypothesis by comparing ten

[1] We do not claim that these operations form a complete or final taxonomy.

Operation	Definition
Abstraction $A : X^n \to C$	The capacity to extract structure or regularities that hold across specific instances, discarding incidental detail in favor of what generalizes Posner & Keele (1968).
Extracts shared features, such as zebras' stripe patterns, while discarding incidental details, such as their habitat.	
Analogy $A : C_1 \times C_2 \to R$	The capacity to map structural relationships from one conceptual domain onto another, supporting reasoning by structural correspondence rather than surface similarity Gentner (1983).
Maps the relation between HORSE and ZEBRA onto DOG and WOLF, a domesticated-to-wild correspondence.	
Categorization $K : X \to C$	The capacity to group instances into classes based on shared properties, enabling stable treatment of perceptually or functionally distinct items as members of the same kind Smith & Medin (1981); Murphy & Medin (1985); Rosch & Mervis (1975).
Assigns a new instance to the ZEBRA category (e.g., a new zebra at the local zoo).	
Communication $M : C \times \mathcal{A} \to C$	The capacity for a concept to be reliably transmitted between agents, supporting shared reference and coordinated use Lazaridou et al. (2017).
Allows the word "zebra" to reliably evoke the same concept in a listener as in the speaker.	
Composition $\mathrm{Comp} : C^n \times R \to C$	The capacity to combine simpler concepts into more complex ones according to systematic rules, such that the meaning of the whole depends predictably on its parts Fodor & Pylyshyn (1988). Note that unlike Analogy, which is about mapping structure *across* domains, Composition is *within* a domain.
Combines the simpler concepts HORSE and STRIPES into ZEBRA, such that the meaning of the whole is predictable from its parts.	
Contextualization $Q : C \times X \to C$	The capacity for a concept's meaning or applicability to shift systematically depending on the surrounding context Barsalou (1982).
Changes its interpretation across uses such as an actual ZEBRA versus a "zebra crossing".	
Discrimination $D : X \times C \times C \to \{0, 1\}$	The capacity to distinguish between concepts that are close or easily confusable, complementary to categorization in that it draws boundaries rather than groupings Green et al. (1966). Note that while Similarity, Categorization, and Discrimination all concern how a concept relates to nearby concepts, they differ in whether the operation groups instances together, draws a boundary between them, or yields a graded judgment of closeness. Note that a system can *categorize* broadly without discriminating finely, or discriminate without producing a full partition.
Distinguishes ZEBRA from the closely confusable concept HORSE.	
Generalization $G : C \times X_{\mathrm{new}} \to \{0, 1\}$	The capacity to extend a concept learned from limited instances to novel instances not previously observed Shepard (1987); Mitchell (1982). Unlike Abstraction, which transforms a set of particular instances into a representation that preserves selected invariant structure across them, Generalization extends an already acquired concept to new instances.
Applies ZEBRA to a previously unseen subspecies, such as the imperial zebra.	
Grounding $R : C \times E \to [0, 1]$	The capacity to connect a concept to something outside the conceptual system itself, such as sensory input, embodied action, or the external world Harnad (1990). Note that unlike Contextualization, which concerns how a concept's meaning shifts *between uses*, Grounding concerns whether a concept connects *at all* to something outside the conceptual system, such as perception or embodied action.
Connects the concept to sensory experience of actual zebras encountered in the world (the way they feel, smell, etc.).	
Inference $I : K \to K$	The capacity to derive new conceptual knowledge from existing knowledge, whether through logical entailment, probabilistic reasoning, or other systematic derivation Osherson et al. (1990). Note that unlike Composition, which builds a new *concept* out of parts, Inference derives a new *belief or conclusion* from existing conceptual knowledge, without necessarily constructing any new composite concept.
Derives that a particular zebra is a herbivore from its membership in ZEBRA, without observing this directly.	
Learning $L : \mathcal{D} \to C$	The capacity to acquire new concepts from experience or data, rather than having them specified in advance Bruner (2017); Valiant (1984). Note that unlike Generalization, which concerns extending an *already-acquired* concept to novel instances, Learning concerns the acquisition of the concept itself.
Acquires the concept ZEBRA in the first place from repeated exposure to instances, for example through picture books or zoo visits.	
Similarity $S : C \times C \to [0, 1]$	The capacity to judge how alike two concepts are, typically supporting graded rather than binary comparison Tversky (1977).
Judging ZEBRA as closer to HORSE than to GIRAFFE.	
Revision $V : C \times E \to C$	The capacity to update or overturn an existing concept when it proves inconsistent with new evidence, as distinct from acquiring an entirely new concept Carey (1985); Chi (1992).
Updates the boundaries of ZEBRA on learning that Okapi, despite a very similar stripe pattern, is not a type of ZEBRA.	

Table 1: The thirteen conceptual operations identified across comparative cognition, cognitive psychology, and AI, used as a representation-independent interface for evaluating mathematical frameworks in Table 2. The order is alphabetical to avoid any claims about importance. Brown text explains what the operation means using the running ZEBRA concept example.

existing mathematical languages using the operations framework introduced above. Appendix B summarizes each framework's assumptions, mathematical objects, constructions, and philosophy.

But mathematical frameworks do not only differ in *how well* they support a given operation; they differ in *what kind of commitment* they make about a concept in the first place. Some languages treat a concept as a self-contained object, defined by its content independent of other concepts, as in a vector or a probability distribution. Others treat a concept as inherently relational, defined by its connections and compositions with other concepts, as in a graph or a logical predicate. Others define concepts by how they emerge, adapt, or stabilize through interaction, as in a dynamical system or a game-theoretic equilibrium. This reflects a prior choice about the basic unit of analysis, which, in turn, constrains the operations a framework can express support.

This distinction organizes the frameworks into three groups: *content frameworks*, which characterize concepts by their standalone properties (§ 4); *relational frameworks*, which characterize concepts by their connections to other concepts (§ 5); and *process frameworks*, which characterize concepts by how they change over time or through interaction (§ 6). Comparing frameworks within each group reveals complementary strengths that share a common commitment (e.g., Geometry and Probability both treat concepts as standalone entities, but emphasize different aspects of their content); comparing across groups reveals deeper tradeoffs. Table 2 provides an overview, and the following sections examine each group and framework in turn.

Operation	Geometry	Probability	Information	Logic	Graphs	Algebra	Category	Dynamical	Optimization	Game-Theory
	Content (§ 4)			*Relational (§ 5)*				*Process (§ 6)*		
Abstraction	○	○	✓	✓	○	✓	✓	×	○	×
Analogy	✓	○	○	○	○	○	✓	×	×	×
Categorization	✓	○	○	○	✓	○	○	○	○	×
Communication	×	×	✓	○	○	×	○	○	×	✓
Composition	○	×	×	✓	○	✓	✓	×	×	×
Contextualization	○	✓	○	×	○	×	○	✓	○	○
Discrimination	✓	○	✓	✓	○	○	○	○	×	×
Generalization	○	✓	○	○	○	○	○	○	✓	×
Grounding	×	✓	○	×	○	×	○	○	○	×
Inference	×	○	×	✓	○	○	○	×	×	○
Learning	✓	✓	✓	×	○	×	○	✓	✓	○
Similarity	✓	○	○	×	○	○	○	○	×	×
Revision	○	✓	○	×	○	○	○	✓	○	○

✓ Naturally supported ○ Supported, but not primary × Not naturally supported

Table 2: Comparison of mathematical frameworks according to the conceptual operations they naturally support, grouped by content, relational, and process frameworks. The operational perspective reveals complementary strengths rather than a single universally superior mathematical language. A framework naturally supports an operation when the operation is represented by a canonical construction with minimal additional representational assumptions. This does not imply that other frameworks cannot implement the operation. See explanation of the rankings in Appendix A. These symbols can be interpreted as: ✓= Primitive; the operation is a canonical object or construction, ○= Induced; the operation can be derived from standard structures, and ×= Added; the operation requires an external representation, objective, or semantics. The classifications indicate how directly each framework supports an operation, not whether the operation is possible within that framework. No operation is exclusive to a single mathematical language; even when a framework does not support an operation natively, it may implement it through additional structures or assumptions.

4 How Are Concepts Represented on Their Own?

Some mathematical languages characterize a concept primarily by what it *is* in isolation. A vector, a probability distribution, or a compressed code all specify a **concept's content on its own terms**, prior to any explicit connection to other concepts. We group these together as *content frameworks*: geometric, probabilistic, and information-theoretic representations. As we show below, each captures a different notion of conceptual content, be it position, uncertainty, or compression, but all treat the concept as a standalone object rather than a node embedded in a network of relations.

4.1 Geometric Representations

Geometric representations, including vector spaces and embedding models, are among the most widely used mathematical languages for representing concepts in modern AI. Their primary strength is expressing concepts through spatial relationships. Similarity, interpolation, clustering, and analogical reasoning emerge naturally from geometric structure, allowing concepts to vary continuously while preserving meaningful neighborhoods. Subspaces further capture shared features and variation, supporting efficient learning and generalization from high-dimensional data.

However, **geometric structure alone does not specify how conceptual representations are composed, grounded, or manipulated in reasoning**. Vector arithmetic captures simple compositional regularities, such as analogy completion, but more complex composition typically requires learned operations or additional structure. Communication, grounding, and inference likewise are not primitive to geometric representations and must be built into the system. Geometry therefore provides a flexible foundation for conceptual relationships and continuous variation, while leaving the mechanisms of structured reasoning underdetermined.

4.2 Probabilistic Representations

Probabilistic frameworks represent concepts through uncertainty. Instead of asking how similar two concepts are, they ask how likely observations, hypotheses, or conceptual assignments are given available evidence. This naturally supports reasoning under uncertainty, evidence accumulation, Bayesian updating, and prediction; operations that are indispensable whenever concepts must be inferred from incomplete or noisy observations, and that probabilistic models handle through principled mechanisms for combining evidence and adapting beliefs as new information arrives.

Their limitations arise elsewhere. **While probability offers an elegant language for uncertainty, it provides comparatively little structure for expressing compositional relationships or higher-order conceptual organization.** Abstraction and symbolic reasoning must therefore be introduced independently of the probabilistic formalism itself.

4.3 Information-Theoretic Representations

Information theory approaches concepts in terms of the information they preserve or discard. Conceptual abstraction becomes a problem of compression: identifying representations that retain behaviorally relevant distinctions while eliminating unnecessary detail.

This perspective naturally supports operations such as abstraction and invariant representation learning, and has become influential in neuroscience and ML, where it provides normative accounts of why intelligent systems develop compressed representations.

Its principal limitation is that **information theory is largely agnostic about the internal organization of conceptual representations**. It can characterize what information is preserved for reasoning, but does not specify the conceptual relationships or operations needed to compose and manipulate that information. Information theory therefore provides powerful principles for optimizing representations without fully specifying the representations and operations that support reasoning.

5 How Do Concepts Relate to Other Concepts?

Other mathematical languages characterize a concept primarily by how it relates to and combines with other concepts. Logical systems, graphs, algebraic structures, and category theory all treat

relations, composition, and structure-preserving mappings as the primary explanatory unit, so that a concept matters mainly through its position within a larger relational system. We group these as *relational frameworks*. As we show below, each emphasizes a different kind of relational structure, be it inference, connectivity, closure, or composition, but all share this shift from content to relation.

5.1 Logical Representations

Logical systems represent concepts through symbolic predicates, relations, and formal rules. Their strength is explicit inference: deduction, consistency checking, explanation, and rule composition support transparent derivation of complex conclusions.

This explicit structure makes logical representations particularly attractive for domains requiring interpretability, formal guarantees, or precise symbolic reasoning. Unlike geometric approaches, logical derivations can be traced step by step, with each inference interpretable on its own terms.

Their principal weakness is that concepts rarely have purely discrete boundaries. **Human concepts are graded, context-dependent, and often ambiguous, properties that classical logic does not capture naturally.** Similarity, continuous variation, and statistical learning therefore require substantial extensions beyond the logical formalism itself.

5.2 Graph-based Representations

Graphs organize concepts through explicit relationships. Rather than emphasizing individual concepts, they focus on the network of connections among them. Hierarchies, dependencies, taxonomies, and semantic networks therefore emerge naturally within graph structures.

This relational perspective supports operations such as navigation, relational reasoning, and structural organization: traversing paths, computing shortest connections, or aggregating over neighborhoods all fall out naturally from graph structure. Knowledge graphs, ontologies, and semantic networks exemplify this approach across AI and cognitive science.

However, **graphs alone provide limited mechanisms for representing graded similarity, uncertainty, or abstraction**. While relationships are explicit, the representations themselves generally require complementary mathematical tools to support graded similarity or probabilistic inference.

5.3 Algebraic Representations

Algebra characterizes concepts through operations and the rules governing their composition, focusing on the structural properties preserved when concepts are combined or transformed. Mathematical objects such as groups, semigroups, lattices, and Boolean algebras provide formal languages for expressing compositional structure, symmetry, and closure.

This perspective naturally supports operations involving composition, decomposition, symmetry, equivalence, and hierarchical organization. Algebraic methods have long influenced symbolic AI, formal concept analysis, and knowledge representation, offering a principled calculus for building complex conceptual structures from simple ones.

At the same time, **algebra typically abstracts away continuous variation, uncertainty, and learning dynamics**. While it provides powerful structural constraints, additional mathematical machinery is often required to capture graded similarity, probabilistic reasoning, or adaptation from data.

5.4 Category-Theoretic Representations

Category theory treats transformations, rather than objects, as primary, asking how concepts relate through structure-preserving mappings. This makes operations such as composition, abstraction, translation between representations, and structural analogy particularly natural. Universal constructions provide principled notions of abstraction, while functors formalize systematic correspondences between representational systems. In contrast to other frameworks, category theory emphasizes interoperability across mathematical structures.

This generality, however, is also its principal limitation. Category theory provides a language for relating structures rather than prescribing the structures themselves. **Operations such as similarity,**

uncertainty, grounding, and learning require additional categorical constructions and domain-specific assumptions to become computationally meaningful. Category theory is therefore well suited to organizing conceptual operations across otherwise disparate formalisms, but does not by itself provide the content- or process-oriented commitments those formalisms capture.

6 How Do Concepts Emerge and Evolve?

A third group of mathematical languages characterizes concepts not by their content or relations, but by how they emerge, adapt, and stabilize over time. Dynamical systems, optimization, and game theory all treat concepts as outcomes of an ongoing process, whether driven by temporal dynamics, adaptation to an objective, or strategic interaction among agents, rather than as fixed content or fixed relational structure. We group these as *process frameworks*. Each captures a different driver of conceptual change, but all share this emphasis on concepts as evolving rather than static.

6.1 Dynamical Systems

Dynamical systems view concepts not as static representations but as states that evolve over time according to underlying dynamics, asking how conceptual states emerge, stabilize, and transform through interaction with their environment. Such models naturally describe attractors, trajectories, bifurcations, and other temporal phenomena that characterize adaptive behavior.

This perspective makes operations such as concept acquisition, conceptual change, and contextual adaptation particularly natural, and extends naturally to memory (as persistent or recurring states) and sequential reasoning (as trajectories through state space). It has therefore become increasingly influential in neuroscience, cognitive science, and ML, where intelligence is often modeled as a continuously evolving process rather than a sequence of discrete computations.

However, **dynamical systems provide comparatively little structure for representing explicit symbolic relationships, compositional reasoning, or semantic hierarchies**. They excel at describing *how* conceptual systems evolve, but less naturally express *what* conceptual structures are.

6.2 Optimization-Based Representations

Optimization approaches concepts as outcomes of objectives optimized under constraints. Concepts emerge because they support efficient prediction, decision making, communication, compression, or other task goals.

This perspective naturally supports learning, adaptation, generalization, and trade-offs between competing objectives. In modern ML, conceptual representations often emerge implicitly from loss functions, inductive biases, and training dynamics.

Its principal limitation is that optimization alone says little about the internal structure of concepts. Similarity, composition, inference, and abstraction depend on the representational framework being optimized, not on optimization itself. Optimization therefore complements, rather than replaces, frameworks for representing and manipulating concepts.

6.3 Game-Theoretic Representations

Game theory studies concepts in the context of strategic interaction among multiple agents. It emphasizes the roles concepts play in collective behavior: how meanings emerge, evolve, and stabilize through communication, cooperation, competition, and negotiation.

This perspective naturally supports operations involving communication, coordination, pragmatic inference, convention formation, and social grounding. It provides a principled framework for understanding how shared conceptual systems arise, adapt, and persist within populations of interacting agents, making it relevant to both human cognition and multi-agent AI.

Its principal limitation is its emphasis on interaction rather than internal representation. Game theory provides limited guidance for representing conceptual structure, similarity, or compositional reasoning within individual agents. It therefore complements representational frameworks by explaining how concepts function and evolve through social interaction.

Theory	Support	Trained	Accuracy
Optimization	○	Yes	**0.833**
Probability	○	Yes	0.798
Geometry	✓	No	**0.607**
Graph	✓	No	0.552
Logic	○	Yes	0.547
Dynamical	○	No	0.541
Information	○	No	0.538
Algebra	○	No	0.394
Category	○	No	0.249
Game-theory	×	–	–

Table 3: Categorization performance across mathematical frameworks. **Support** indicates the framework's theoretical support for categorization (Table 2).

7 Categorization Test-Case

To better illustrate how different mathematical frameworks answer questions about concept operations, we ask each framework the same question: *how would this concept be categorized?*. We use a human categorization judgments dataset from Shani et al. (2026). Categorization is a useful test case because almost every framework has something to say about it, yet they view the operation differently. We explore nine frameworks, excluding game theory, which does not naturally support categorization (Table 2). For each, we implement the corresponding procedure on the same items and compare its output against the same human judgments.

Operationalizations. *Geometry* treats a category as a region and clusters by Euclidean distance. *Probability* models each category as a class-conditional density and assigns by posterior. *Optimization* treats categorization as empirical risk minimization and fits a discriminative probe. *Information* uses the Information Bottleneck (Tishby et al., 2000), compressing items while preserving a relevance variable. *Logic* defines a category by features necessary and jointly sufficient for membership. *Algebra* builds the Galois concept lattice over items and their features (Ganter et al., 1999). *Graph* clusters by shortest-path distance in a knowledge graph; *Structural* clusters by structural equivalence in that graph, discarding node identity. *Dynamical* treats a category as an attractor and measures how stable an item's representation is across training checkpoints. Table 5 summarizes the operationalizations and results.

Protocol. Every method returns a partition of the items, scored against the human partition. When a method admits a choice of cluster number, we set it to the number of human categories, preventing any method from benefiting from its own granularity. Methods that require fitting use nested cross-validation grouped by item, so no item appears in both training and test folds. We report Adjusted Rand Index and Adjusted Mutual Information, both corrected for chance agreement, alongside uncorrected NMI, with confidence intervals obtained by subsampling items, and Accuracy for trained methods with Matched Accuracy for all methods (on the same held-out set). Matched accuracy is cluster accuracy after optimal Hungarian matching to the human categories. Each procedure is tested against a null matched to its confounds: initialization for clustering, checkpoint order for trajectories, and information-budget variation for relevance comparisons. Full methodological details and additional metrics are provided in Appendix B.1.

7.1 Quantitative interpretation

The methods differ in ways that cannot be fully equalized. Three of the nine are supervised, so their predictive scores are not directly comparable to the six unsupervised methods. The methods also require different auxiliary resources and therefore cover different subsets of the items: the knowledge graph covers nearly all items, whereas property norms cover 60-76%. Finally, accuracy is defined only for the supervised methods; for the others, we report cluster accuracy after optimal matching, which measures agreement with the human partition under a different criterion. We therefore do

not interpret Table 5 as a ranking, and acknowledge that because these frameworks speak different "languages," comparisons is imperfect. We thus take this empirical analysis with a grain of salt.

Having said that, these scores can still indicate how well each operationalization recovers the human category structure under its particular mathematical commitments and input representation. A low score is evidence that the corresponding operationalization provides a poor account of human categorization in this setting, although it does not by itself identify whether the limitation lies in the mathematical framework, its representation, its auxiliary data, or its implementation. Conversely, higher recovery provides evidence that the framework captures aspects of categorization reflected in the human judgments.

The two highest-performing methods are supervised, consistent with the advantage of access to category labels during fitting. Among the non-supervised methods, the strongest performers are those classified as ✓ in Table 2, suggesting that frameworks that naturally support categorization also tend to provide stronger accounts of human category structure when instantiated on this task.

7.2 Qualitative interpretation

We next examine *crowbar*, a Carpenter's Tool in the human data and the item on which the nine mathematical implementations disagree most. It receives seven distinct majority categories, and the average overlap between the sets of items with which it is grouped is only 0.14 (see Appendix B.2). The four embedding-based accounts place it among birds, the property-norm accounts place it among weapons, and the knowledge-graph account assigns it no coherent neighborhood.

The embedding-based error has a sub-lexical origin. Under mean pooling, $\cos(\textit{crowbar}, \textit{crow}) = 0.735$, and nine of its ten nearest neighbors are birds, while its similarity to other tools does not exceed 0.32. The compound therefore inherits the neighborhood of its leading subword, an effect also visible in pairs such as *sim*(blackberry, blackjack)=0.719. This example also reveals a limitation of apparent convergence across methods. The four embedding-based accounts agree because they share the same input representation and its associated bias, so their agreement is not four independent pieces of evidence. Conversely, disagreement across methods using different representations can reveal which source of evidence drives an error, something an aggregate score obscures.

8 Conclusions & Future Work

We have argued that a mathematics of concepts should be organized around the operations concepts support, such as similarity, composition, generalization, grounding, and revision, rather than around a mathematical framework. Comparing ten mathematical frameworks through this lens shows that each makes different conceptual operations natural, reflecting different commitments to conceptual content, relational structure, and process. No framework in our comparison naturally supports the full set of operations, suggesting that **a mathematics of concepts will likely require combining complementary mathematical frameworks rather than refining a single one**.

Our categorization experiment provides an initial empirical test of this framework. The results suggest that the theoretical notion of *natural support* is not merely descriptive: among non-supervised methods, the frameworks classified as naturally supporting categorization achieve the strongest recovery of human category structure. At the same time, supervised methods can achieve higher performance by introducing information that is not native to the framework. The results therefore suggest a useful distinction between *native mathematical support* and *performance through added machinery*: a framework's mathematical commitments constrain what it can express naturally, while additional supervision, representations, and auxiliary data can extend what it achieves in practice.

Several directions follow from this framing. First, the empirical analysis should be extended beyond categorization to the broader set of conceptual operations, testing whether the theoretical distinctions in Table 2 predict behavior across tasks and representations. Second, the comparison should be expanded to additional mathematical languages and hybrid formalisms, as well as to a broader and empirically validated taxonomy of conceptual operations. Most importantly, the framework opens a constructive question: **can we develop mathematical formalisms that support a broad, human-like range of conceptual operations?** We encourage the field to build a mathematical language that makes concepts' full range of operations explicit.

AI USE STATEMENT

LLMs were used to locate relevant literature, edit language, discuss mathematical frameworks, and implement the empirical categorization experiment.

ETHICS STATEMENT

The authors of this submission do not see any potential violation of the ICLR Code of Ethics.

REPRODUCIBILITY STATEMENT

We release the code here: `https://anonymous.4open.science/r/concept_math-B3BA/README.md`. The data used for the empirical analysis of categorization was published by Shani et al. (2026) and can be accessed here: `https://huggingface.co/datasets/CShani/human-concepts`.

REFERENCES

Steve Awodey. *Category theory*, volume 49. Oxford university press, 2006.

Franz Baader and Tobias Nipkow. *Term rewriting and all that*, volume 1. Cambridge university press Cambridge, 1998.

Lawrence W Barsalou. Context-independent and context-dependent information in concepts. *Memory & cognition*, 10(1):82–93, 1982.

Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2(3):5, 2018.

Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: Structure and dynamics. *Physics reports*, 874:1–92, 2020.

Randall D Beer. Dynamical approaches to cognitive science. *Trends in cognitive sciences*, 4(3): 91–99, 2000.

James O Berger. Statistical decision theory. In *The New Palgrave Dictionary of Economics*, pp. 1–6. Springer, 1987.

Patrick Billingsley. *Probability and measure: anniversary edition*, volume 446. wiley Hoboken, NJ, 2012.

Karol Borsuk. *Foundations of geometry*. Courier Dover Publications, 2018.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Jerome Bruner. *A study of thinking*. Routledge, 2017.

Susan. Carey. *Conceptual change in childhood*. MIT Press series in learning, development and conceptual change. MIT Press, Cambridge, Mass, 1985.

George Casella and Roger Berger. *Statistical inference*. CRC press, 2024.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Michelene Chi. Conceptual change within and across ontological categories: Examples from learning and discovery in science. 1992.

Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. Mathematical foundations for a compositional distributional model of meaning. *arXiv preprint arXiv:1003.4394*, 2010.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

P. M. Cohn. Universal algebra, January 2003. URL https://ideas.repec.org/h/spr/sprchp/978-1-4471-0039-3_1.html.

John Collins. The nature of linguistic variables. In *The Oxford Handbook of Topics in Philosophy*. Oxford University Press. ISBN 9780199935314. doi: 10.1093/oxfordhb/9780199935314.013.004. URL https://doi.org/10.1093/oxfordhb/9780199935314.013.004.

Bruno Courcelle. The monadic second-order logic of graphs vii: graphs as relational structures. *Theoretical Computer Science*, 101(1):3–33, 1992. ISSN 0304-3975. doi: https://doi.org/10.1016/0304-3975(92)90148-9. URL https://www.sciencedirect.com/science/article/pii/0304397592901489.

Thomas M Cover, Joy A Thomas, and John Kieffer. *Elements of information theory*, volume 2. wiley New York, 1991.

Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pp. 1431–1451, 1982.

Luc Devroye, László Györfi, and Gábor Lugosi. *A probabilistic theory of pattern recognition*, volume 31. springer New York, 1996.

Manfredo P Do Carmo. *Differential geometry of curves and surfaces: revised and updated second edition*. Courier Dover Publications, 2016.

Hartmut Ehrig, Karsten Ehrig, Ulrike Prange, and Gabriele Taentzer. *Fundamentals of algebraic graph transformation*. Springer, 2006.

Herbert B Enderton. *A mathematical introduction to logic*. Elsevier, 2001.

Ronald Fagin, Joseph Y Halpern, Yoram Moses, and Moshe Vardi. *Reasoning about knowledge*. The MIT press, 1995.

Jerry A Fodor and Zenon W Pylyshyn. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2):3–71, 1988.

Santo Fortunato et al. Community detection in graphs. 2010.

Drew Fudenberg and Jean Tirole. *Game theory*. MIT press, 1991.

Bernhard Ganter, Rudolf Wille, and Cornelia Franzke. *Formal concept analysis: mathematical foundations*, volume 284. Springer, 1999.

Peter Gärdenfors. *Conceptual spaces*, volume 3. MIT press Cambridge, MA, 2000.

Dedre Gentner. Structure-mapping: A theoretical framework for analogy. *Cognitive science*, 7(2): 155–170, 1983.

Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553): 452–459, 2015.

Alison L Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002.

David Marvin Green, John A Swets, et al. *Signal detection theory and psychophysics*, volume 1. Wiley New York, 1966.

Thomas L Griffiths, Adam N Sanborn, Kevin R Canini, and Daniel J Navarro. Categorization as nonparametric bayesian density estimation. *The probabilistic mind: Prospects for Bayesian cognitive science*, pp. 303–328, 2008.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Stevan Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346, 1990.

Peter E Hart, David G Stork, and RO Duda. *Pattern classification*. Wiley Hoboken, 2001.

Gerhard Heinzmann and Jean Petitot. The functional role of structures in bourbaki. In Erich H. Reck and Georg Schiemer (eds.), *The Prehistory of Mathematical Structuralism*. Oxford University Press, 04 2020. ISBN 9780190641221. doi: 10.1093/oso/9780190641221.003.0008. URL https://doi.org/10.1093/oso/9780190641221.003.0008.

Pavol Hell, Jaroslav Neset, et al. *Graphs and homomorphisms*, volume 34. Oxford University Press, 2026.

Marvin Hofer, Daniel Obraczka, Alieh Saeedi, Hanna Köpcke, and Erhard Rahm. Construction of knowledge graphs: Current state and challenges. *Information*, 15(8):509, 2024.

Edwin T Jaynes and G Larry Bretthorst. *Probability theory: The logic of science*, volume 1. Cambridge university press Cambridge, 2003.

Edwin Thompson Jaynes. *Probability theory: the logic of science*. Washington University, 1996.

W Jeffrey Johnston and Stefano Fusi. Abstract representations emerge naturally in neural networks trained to perform multiple tasks. *Nature Communications*, 14(1):1040, 2023.

Robinson Derek JS. *Abstract algebra*, volume 1. De Gruyter, 2022.

Olav Kallenberg. *Probabilistic symmetries and invariance principles*. Springer, 2005.

Daphne Koller and Nir Friedman. *Probabilistic graphical models*, volume 165. MIT press Cambridge, 2009.

Pravesh Kothari, Raghu Meka, and Prasad Raghavendra. Theory beyond theorems and proofs: A guest post. Shtetl-Optimized (blog), 2026. URL https://scottaaronson.blog/?p=10116. Guest post hosted by Scott Aaronson.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

Stephen Lack and Paweł Sobociński. Adhesive and quasiadhesive categories. *RAIRO-Theoretical Informatics and Applications*, 39(3):511–545, 2005.

Leon Lang LawrenceC and Erik Jenner. Natural abstractions: Key claims, theorems, and critiques. 2023.

Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. Multi-agent cooperation and the emergence of (natural) language. In *ICLR*, 2017.

Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.

Francois Lorrain and Harrison C White. Structural equivalence of individuals in social networks. *The Journal of mathematical sociology*, 1(1):49–80, 1971.

Saunders Mac Lane and Saunders MacLane. *Categories for the working mathematician*, volume 2. Springer, 1971.

George E Martin. *Transformation geometry: An introduction to symmetry*. Springer Science & Business Media, 2012.

Boris Mirkin. *Mathematical classification and clustering*. Springer Science & Business Media, 2013.

Tom M Mitchell. Generalization as search. *Artificial intelligence*, 18(2):203–226, 1982.

Stephen Muggleton. Inductive logic programming. *New generation computing*, 8(4):295–318, 1991.

Gregory L Murphy and Douglas L Medin. The role of theories in conceptual coherence. *Psychological review*, 92(3):289, 1985.

Mark Newman. *Networks: An Introduction*. Oxford University Press, 03 2010. ISBN 9780199206650. doi: 10.1093/acprof:oso/9780199206650.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199206650.001.0001.

Martin J Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994.

Daniel N Osherson, Edward E Smith, Ormond Wilkie, Alejandro Lopez, and Eldar Shafir. Category-based induction. *Psychological review*, 97(2):185, 1990.

Judea Pearl. *Probabilistic reasoning in intelligent systems*, volume 88. Elsevier, 2014.

Michael I Posner and Steven W Keele. On the genesis of abstract ideas. *Journal of experimental psychology*, 77(3p1):353, 1968.

Willard Van Orman Quine. *Methods of logic*. Harvard University Press, 1982.

Eleanor Rosch and Carolyn B Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4):573–605, 1975.

Grzegorz Rozenberg. *Handbook of graph grammars and computing by graph transformation*, volume 1. World scientific, 1997.

Chen Shani, Liron Soffer, Dan Jurafsky, Yann LeCun, and Ravid Shwartz-Ziv. From tokens to thoughts: How llms and humans trade compression for meaning. In *International Conference on Learning Representations*, volume 2026, pp. 83019–83055, 2026.

Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

Roger N Shepard. Toward a universal law of generalization for psychological science. *Science*, 237 (4820):1317–1323, 1987.

Edward E Smith and Douglas L Medin. *Categories and concepts*. Harvard University Press, 1981.

John Stillwell, KA Ribet, and S.. Axler. *The four pillars of geometry*, volume 2. Springer, 2005.

Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC press, 2024.

Richard S Sutton, Andrew G Barto, and Andrew Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

Andrzej Tarlecki, Rod M. Burstall, and Joseph A. Goguen. Some fundamental algebraic tools for the semantics of computation: Part 3. indexed categories. *Theoretical Computer Science*, 91(2):239–264, 1991. ISSN 0304-3975. doi: https://doi.org/10.1016/0304-3975(91)90085-G. URL https://www.sciencedirect.com/science/article/pii/030439759190085G.

Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. *arXiv preprint physics/0004057*, 2000.

Amos Tversky. Features of similarity. *Psychological review*, 84(4):327, 1977.

Marius Usher. A statistical referential theory of content: Using information theory to account for misrepresentation. *Mind & Language*, 16(3):311–334, 2001.

Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

Lieven Vandenberghe and Stephen Boyd. *Convex optimization*, volume 1. Cambridge university press Cambridge, 2004.

Vladimir Vapnik and Rauf Izmailov. Rethinking statistical learning theory: learning using statistical invariants. *Machine Learning*, 108(3):381–423, 2019.

Vladimir Naumovich Vapnik. *Statistical learning theory* {*Adaptive and learning systems for signal processing, communications, and control*}. Wiley and Sons, 1998.

Douglas Brent West et al. *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River, 2001.

Rose Yu and Rui Wang. Learning dynamical systems from data: An introduction to physics-guided deep learning. *Proceedings of the National Academy of Sciences*, 121(27):e2311808121, 2024.

A Rationale for Table 2

The assignments in Table 2 are based on whether a mathematical framework provides structures or operations that naturally correspond to each conceptual operation. We use three levels of support: *naturally supported* (✓) when the operation corresponds to a canonical object or operation of the framework; *supported, but not primary* (○) when the framework provides relevant mathematical machinery but the operation is not a central object of the framework; and *not naturally supported* (×) when expressing the operation requires substantial additional assumptions or machinery external to the framework. The rationale for each operation is given below.

Abstraction. Abstraction concerns extracting structure that is invariant across particular instances while discarding incidental detail (Posner & Keele, 1968). Information theory (✓) naturally supports this through representations that retain relevant information while discarding irrelevant variation (Cover et al., 1991). Algebra (✓) and category theory (✓) provide particularly direct structural forms of abstraction: algebra identifies common structure through abstract operations and axioms (Cohn, 2003; Heinzmann & Petitot, 2020), while category theory abstracts away from the internal details of objects in favor of their relationships and mappings (Mac Lane & MacLane, 1971). Logic (✓) also supports abstraction through variables, predicates, quantification, and formal structures that separate general properties from particular instances (Quine, 1982; Collins). Geometry (○) can support abstraction by identifying invariant geometric structure under transformations (Martin, 2012). Probability (○) can likewise support abstraction by identifying probabilistic structure that remains invariant under transformations or symmetries (Kallenberg, 2005). Graphs (○) provide abstraction by representing entities and their relations independently of their concrete implementation, and by supporting structural representations that preserve relational patterns while discarding incidental detail (Battiston et al., 2020), while optimization (○) supports abstraction more indirectly by selecting representations or solutions that preserve task-relevant structure under an explicitly specified objective (Tishby et al., 2000). Dynamical systems (×) and game-theory (×) do not naturally support abstraction.

Analogy. Analogy concerns mapping structural relationships from one conceptual domain onto another (Gentner, 1983). Geometry (✓) naturally supports analogy through transformations, isometries, correspondences, and mappings between geometric structures (Martin, 2012). Category theory (✓) provides an especially direct formal language for structural correspondence through functors and related structure-preserving mappings (Osherson et al., 1990). Logic (○), graphs (○), and algebra (○) can express structural correspondences between formal systems through interpretations, homomorphisms, and other structure-preserving mappings (Enderton, 2001; Hell et al., 2026; Cohn, 2003). Probability (○) and information theory (○) can support analogy through probabilistic or information-theoretic comparisons between domains (Gibbs & Su, 2002; Cover et al., 1991), but these are generally secondary uses. Dynamical systems (×), optimization (×), and game theory (×) do not naturally provide cross-domain structural analogy.

Categorization. Categorization concerns grouping distinct instances into classes on the basis of shared properties (Smith & Medin, 1981; Murphy & Medin, 1985; Rosch & Mervis, 1975). Geometry (✓) naturally supports categorization through clustering, regions, neighborhoods, and geometric

structures (Mirkin, 2013), and graphs (✓) support categorization when categories are represented as subgraphs, clusters, or communities (Martin, 2012; Fortunato et al., 2010; Hell et al., 2026) Probability (○) supports probabilistic class membership and distributions over instances (Griffiths et al., 2008); Information theory (○) can characterize category structure through mutual information shared by instances and category variables (Cover et al., 1991). Logic (○) provides predicates that can define classes and formal distinctions, while equivalence relations provide a canonical mathematical mechanism for partitioning objects into classes, although categorization is not itself a central logical operation (Enderton, 2001). Algebra (○) provides equivalence relations, quotient structures, and subalgebras that can partition objects into classes (Cohn, 2003); Category theory (○) organizes objects through morphisms and structural relationships, allowing objects with shared structural properties to be grouped or identified up to isomorphism (Mac Lane & MacLane, 1971). Dynamical systems (○) can partition state spaces into regions with qualitatively similar trajectories or asymptotic behavior (Strogatz, 2024). Optimization (○) can support categorization when category membership is formulated as an optimization or decision problem, but this requires an additional objective or loss function (Hart et al., 2001). Game theory (×) does not naturally provide categorization as a mathematical primitive.

Communication. Communication concerns the reliable transmission of a concept between agents (Lazaridou et al., 2017). Information theory (✓) provides a way to measure information lose during communication (Cover et al., 1991). Game theory (✓) provides canonical models of communication through signaling games, cheap-talk models, and strategic information transmission (Crawford & Sobel, 1982). Category theory (○) can support formal accounts of communication by representing compositional mappings between structured representational systems, although the communicative interpretation requires additional assumptions about encoding and interpretation (Coecke et al., 2010). Dynamical systems (○) can be adapted such that each system represents an agent's conceptual representation and communication alters the system (Strogatz, 2024). Logic (○) can represent formal messages, communication protocols, and agents' knowledge states (Fagin et al., 1995), while graphs (○) can represent communication networks by modeling agents as nodes and communication links as edges, with paths capturing possible routes of information flow (Newman, 2010). These provide useful machinery, but communication is not itself a primary operation of logic or graph theory. Geometry (×), probability (×), and algebra (×) do not naturally treats inter-agent conceptual transmission as a central operation.

Composition. Composition concerns combining simpler concepts into a structured whole according to systematic rules (Fodor & Pylyshyn, 1988). Logic (✓) provides canonical compositional operations through logical connectives and the recursive construction of well-formed expressions (Quine, 1982). Algebra (✓) provides composition through binary operations and algebraic structures (Cohn, 2003), while category theory (✓) makes composition of morphisms a foundational operation (Mac Lane & MacLane, 1971). Graphs (○) support composition through graph products, substitution, and other graph-combination operations, but composition is not as fundamental as it is in algebra or category theory (Hell et al., 2026). Geometry (○) can represent composite structures through geometric construction and composition of simpler configurations, but composition is secondary to its focus on geometric structure and invariance (Borsuk, 2018). Probability (×), information theory (×), dynamical systems (×), optimization (×), and game theory (×) can be used in compositional models but do not themselves make conceptual composition a primary mathematical operation.

Contextualization. Contextualization concerns systematic changes in the meaning or applicability of a concept as surrounding circumstances change (Barsalou, 1982). Probability (✓) naturally supports contextual dependence through conditional distributions and conditional probabilities such as Markov decision processes (Jaynes & Bretthorst, 2003). Dynamical systems (✓) provide a direct representation of state-dependent behavior, in which the current state changes the system's subsequent evolution (Strogatz, 2024). Game theory (○) represents how an agent's available actions and payoffs depend on the strategic context created by other agents, although contextualized meaning is not its primary object (Osborne & Rubinstein, 1994). Information theory (○) can represent contextual dependence directly through conditional entropy, conditional mutual information, and related conditional measures (Cover et al., 1991). Other frameworks support contextual dependence through additional structural constructions: geometry (○) through context-dependent metrics or con-

straints(Stillwell et al., 2005), graphs (○) through context-dependent relational structure (Courcelle, 1992), category theory (○) through indexed or fibered constructions (Tarlecki et al., 1991), and optimization (○) through context-dependent objectives or constraints (Vandenberghe & Boyd, 2004). In each case, however, contextual dependence is introduced through an additional construction rather than being a primitive organizing principle of the framework. Logic does not naturally provide context-sensitive conceptual meaning without extending the formalism with an explicit representation of context. Logic (×) and algebra (×) do not inherently represent contextual dependence.

Discrimination. Discrimination concerns distinguishing between concepts that are close or easily confusable (Green et al., 1966). Geometry (✓) naturally supports discrimination through distances, boundaries, and separations in a metric or geometric space (Stillwell et al., 2005). Logic (✓)provides canonical distinctions through predicates, truth conditions, and entailment (Enderton, 2001). Information theory (✓) provides measures of statistical distinguishability using measures such as Kullback-Leibler and Jensen-Shannon divergence (Kullback & Leibler, 1951; Lin, 1991). Probability (○) supports discrimination through posterior probabilities and likelihood ratios, which quantify relative support for competing hypotheses (Berger, 1987; Casella & Berger, 2024). Graphs (○), algebra (○), category theory (○), and dynamical systems (○) can distinguish entities through structural properties, such as graph connectivity, algebraic relations, categorical morphisms and isomorphisms, or differences in dynamical states and trajectories, but they do not intrinsically provide a criterion for deciding which distinctions are relevant for discrimination (West et al., 2001; JS, 2022; Mac Lane & MacLane, 1971; Strogatz, 2024). Optimization (×) and game theory (×) do not intrinsically provide mechanisms for discrimination; optimization selects solutions according to a specified objective, while game theory characterizes strategic interactions according to specified preferences, actions, and payoffs.

Generalization. Generalization concerns extending a concept acquired from limited observations to novel instances (Shepard, 1987; Mitchell, 1982). Probability (✓) provides canonical machinery for generalization through statistical inference and prediction to unobserved or future cases (Casella & Berger, 2024), while optimization (✓) provides the objective-based machinery underlying many formal treatments of learning, including the minimization of empirical risk used to study generalization (Vapnik & Izmailov, 2019). Geometry (○) can support generalization through metric neighborhoods, spatial smoothness, interpolation, and geometric decision regions, but these mechanisms do not by themselves specify how experience determines behavior on unseen instances (Devroye et al., 1996). Information theory (○) provides tools for characterizing the relationship between learned representations and generalization, but does not itself define a generalization procedure (Cover et al., 1991). Logic (○) can support generalization by inducing rules from examples; graphs (○) by learning reusable relational or local structural patterns; algebra (○) through reusable algebraic laws and invariants; category theory (○) through structure-preserving mappings; and dynamical systems (○) by learning dynamical laws that predict unobserved states or trajectories. In each case, however, an additional learning or inference mechanism is required to determine which structures transfer to unseen instances (Hamilton et al., 2017; Yu & Wang, 2024; Muggleton, 1991; Cohn, 2003; Mac Lane & MacLane, 1971). Game theory (×) does not support generalization from one solution to another.

Grounding. Grounding concerns connecting conceptual representations to something external to the conceptual system, such as sensory observations, actions, or the physical environment (Harnad, 1990). Probability (✓) naturally supports grounding by linking latent variables or hypotheses to observable evidence through generative and observation models (Pearl, 2014). Graphs (○) can support grounding by representing relations between abstract concepts and entities or observations in the world, as in knowledge graphs and scene graphs, but grounding requires an additional mapping between graph representations and external or perceptual entities (Hofer et al., 2024). Information theory (○) can characterize the statistical relationship between conceptual representations and their external referents through measures such as mutual information, but statistical dependence alone does not establish a grounding or reference relation (Usher, 2001). Category theory (○) can represent mappings between conceptual and external domains through morphisms and functors, although the identification and interpretation of those domains must be supplied separately (Awodey, 2006; Mac Lane & MacLane, 1971). Dynamical systems (○) can model coupling between internal states and environmental variables through coupled dynamical equations, although such coupling does not by itself establish a grounding relation (Beer, 2000). Optimization (○) can connect internal repre-

sentations to actions and environmental outcomes by optimizing task objectives defined over their consequences (Sutton et al., 1998). Algebra (×), logic (×), geometry (×), and game theory (×) do not intrinsically provide an external grounding relation. Game theory (×) can represent interactions with environments or other agents, but interaction alone does not establish semantic grounding.

Inference. Inference concerns deriving new beliefs or conclusions from existing conceptual knowledge (Osherson et al., 1990). Logic (✓) provides canonical machinery for inference through formal deduction and logical entailment (Enderton, 2001). Probability (○) supports probabilistic inference through conditional probability, Bayesian inference, and probabilistic graphical models, but inference depends on a specified probabilistic model and inferential procedure (Pearl, 2014). Graphs (○) support inference when relational structure encodes dependencies or constraints, allowing information to propagate through paths and connected variables (Koller & Friedman, 2009). Algebra (○) supports reasoning through algebraic laws, identities, and structure-preserving transformations, while category theory (○) supports compositional reasoning through morphism composition, commutative diagrams, and universal properties (JS, 2022; Awodey, 2006). Game theory (○) supports inference about strategic behavior through equilibrium and related solution concepts (Fudenberg & Tirole, 1991). Geometry (×) and dynamical systems (×) can support inference about spatial or temporal systems, but it is a different type of inference from the intended one, and optimization (×) primarily specifies the selection of solutions under an objective and therefore does not, by itself, constitute conceptual inference.

Learning. Learning concerns the acquisition of concepts from experience or data (Bruner, 2017; Valiant, 1984). Probability (✓) provides a canonical framework for learning through statistical inference and posterior updates (Ghahramani, 2015), and optimization (✓) provides the central machinery for parameter estimation and empirical risk minimization (Vapnik, 1998). Information theory (✓) naturally characterizes learning in terms of information acquisition and better compression (Shannon, 1948). Geometry (✓) can support learning through geometric models of representations and decision boundaries (Bronstein et al., 2017). Dynamical systems (✓) naturally represent learning when it is formulated as a trajectory through parameter or state space (Chen et al., 2018). Graphs (○) and category theory (○) can support learning through graph-structured or relational representations, but learning requires additional algorithms or objectives (Battaglia et al., 2018; Coecke et al., 2010). Logic (×) and algebra (×) do not naturally provide mechanisms for acquiring concepts from data.

Similarity. Similarity concerns the graded comparison of concepts according to how alike they are (Tversky, 1977). Geometry (✓) provides a canonical account of similarity through metrics, distances, angles, and transformations, and is therefore naturally suited to this operation (Do Carmo, 2016; Stillwell et al., 2005), while information theory (○) provides closely related measures of distinguishability such as Kullback-Leibler and Jensen-Shannon divergence (Cover et al., 1991). Probability (○) supports graded similarity through likelihoods and distances or divergences between probability distributions (Billingsley, 2012); graphs (○) can support similarity through graph distances, shared neighborhoods, or structural equivalence (Lorrain & White, 1971); algebra (○) through relations or metrics defined on algebraic structures (Cohen & Welling, 2016); category theory (○) through structure-preserving mappings (Mac Lane & MacLane, 1971); and dynamical systems (○) through distances between states or trajectories (Shannon, 1948). In each case, however, similarity is derived from the framework's underlying structure rather than being a primary object of the framework (). Logic (×), optimization (×), and game theory (×) do not naturally provide graded measures of conceptual similarity.

Revision. Revision concerns updating or overturning an existing concept in response to new evidence (Carey, 1985; Chi, 1992). Probability (✓) provides the most direct canonical mechanism through Bayesian updating, in which beliefs are systematically revised in light of observations (Jaynes, 1996). Dynamical systems (✓) naturally represent revision as state evolution, although the interpretation of the changing state as a revised concept must be supplied (Shannon, 1948). Optimization (○) can support revision through iterative parameter updates, but this depends on an externally specified objective (Sutton et al., 1998). Graphs (○), algebra (○), and category theory (○) can represent the result of a revision, e.g., an edge or node update in a graph, an algebraic rewriting rule, a new morphism in a category (Ehrig et al., 2006; Rozenberg, 1997; Baader & Nipkow, 1998;

Lack & Sobociński, 2005). However, none specifies when incoming evidence should trigger such a change or which of several possible transformations counts as the correct revision, without an externally supplied revision rule. Geometry (○) can represent movement or deformation of representations, but does not itself provide a mechanism for deciding when an existing concept should be revised (Gärdenfors, 2000). Logic (×) can represent nonmonotonic or belief-revision systems, but classical logic alone does not naturally support revision in response to contradictory evidence.

Together, these assessments define the operational comparison in Table 2. The assignments should not be interpreted as claims about the expressive limits of any framework. Rather, they identify which conceptual operations each mathematical perspective makes most natural. This distinction is important: a framework may be capable of representing an operation without making that operation mathematically primitive. The resulting complementarity suggests that a general mathematics of concepts may be better understood as an integration of mathematical perspectives, each providing different portions of the conceptual interface, than as the search for a single universally sufficient formalism.

B Mathematical Frameworks

Table 4 provides a concise overview of the ten mathematical frameworks considered in our analysis, summarizing their core assumptions, mathematical objects, canonical constructions, and guiding philosophies. We treat these frameworks as distinct mathematical languages for describing conceptual structure: each makes certain properties and operations explicit while leaving others implicit or requiring additional assumptions. The frameworks therefore differ not in whether they can ultimately implement a given conceptual operation, but in how naturally that operation is expressed within their native mathematical structure.

B.1 Empirical Analysis of Categorization

Table 5 depicts the results of the empirical investigation of the categorization operation of concepts.

Three of the nine are supervised and use category labels during fitting, so their predictive scores are not directly comparable to the six unsupervised methods. The methods also require different auxiliary resources and therefore cover different subsets of the items: the knowledge graph covers nearly all items, whereas property norms cover 60-76%. Finally, predictive accuracy is defined only for the supervised methods; for the others, we report cluster accuracy after optimal matching, which measures agreement with the human partition under a different criterion. We therefore do not interpret Table 5 as a ranking.

These scores indicate how well each operationalization recovers the human category structure under its particular mathematical commitments and input representation. Thus, a low score is evidence that the corresponding operationalization provides a poor account of human categorization in this setting, although it does not by itself identify whether the limitation lies in the mathematical framework, its representation, its auxiliary data, or its implementation. Conversely, higher recovery provides evidence that the framework captures aspects of categorization reflected in the human judgments.

The two highest-performing methods are supervised, consistent with the advantage of access to category labels during fitting. Among the non-supervised methods, the strongest performers are those classified as ✓ in Table 2, suggesting that frameworks that naturally support categorization also tend to provide stronger accounts of human category structure when instantiated on this task.

B.2 Qualitative Analysis of *crowbar*

Table 6 provides the full breakdown of the categorizations assigned to *crowbar* by the nine mathematical implementations. In the human data, *crowbar* belongs to the category Carpenter's Tool, but the implementations assign it to seven distinct majority categories. The average overlap between the sets of items with which *crowbar* is grouped is only 0.14, making it the item with the greatest disagreement across methods.

The disagreement also reveals distinct sources of error. The four embedding-based accounts place *crowbar* among birds. Under mean pooling, $\cos(\textit{crowbar}, \textit{crow}) = 0.735$, and nine of its ten nearest

neighbors are birds, while its similarity to other tools does not exceed 0.32. The compound therefore inherits the neighborhood of its leading subword. A similar effect appears in pairs such as *blackberry-blackjack*, which have cosine similarity 0.719. In contrast, the property-norm accounts place *crowbar* among weapons, while the knowledge-graph account assigns it no coherent neighborhood.

This example illustrates why agreement across implementations should not automatically be interpreted as independent evidence. The four embedding-based accounts share the same input representation and therefore can inherit the same representational bias. Conversely, disagreement between methods based on different representations can help identify which source of information drives an error. Thus, the qualitative analysis complements the aggregate scores by revealing the representational assumptions underlying both convergence and divergence across mathematical accounts.

Framework	Core assumption	Mathematical objects	Canonical constructions	Guiding philosophy
Geometry	Concepts occupy structured spaces in which spatial relationships encode meaningful relationships.	Points, vectors, spaces, distances, subspaces	Distance, similarity, projection, interpolation, vector arithmetic, transformations	Understand concepts through their position and structure in a space: what matters is where concepts lie, how far apart they are, and how they can be transformed or combined geometrically.
Probability	Conceptual knowledge is inherently uncertain, and representations can be understood as distributions over possible states or interpretations.	Random variables, probability distributions, conditional distributions	Conditioning, marginalization, Bayesian updating, expectation, sampling	Represent what is known not as a fixed value but as a distribution over possibilities, and update that distribution as evidence arrives.
Information	Representations differ in how much relevant information they preserve, discard, or share.	Entropy, mutual information, information measures	Compression, information bottlenecks, information maximization, rate–distortion	Characterize a good representation by the information it retains about what matters while discarding irrelevant detail.
Logic	Conceptual knowledge can be expressed through explicit propositions and rules whose validity is determined by formal structure.	Symbols, predicates, propositions, rules, logical structures	Deduction, implication, substitution, unification, rule application, logical composition	Make conceptual structure explicit and derive consequences through formally specified rules, rather than relying on graded similarity alone.
Graphs	Concepts are entities whose meaning and behavior depend on explicit relationships to other entities.	Nodes, edges, paths, graphs, networks	Traversal, connectivity, path finding, neighborhood operations, graph composition	Understand a conceptual system through its network of entities and relations: what matters is which concepts are connected, how, and through what paths.
Algebra	Conceptual systems can be characterized by the operations and transformations that act on their elements.	Elements, operators, groups, algebras, transformations	Composition, inversion, transformation, equivalence, homomorphism	Understand concepts through what can be done with them and how operations compose, rather than primarily through their absolute representation.
Category	The structure of a system is determined by objects and the mappings between them, with emphasis on relationships and compositionality.	Objects, morphisms, categories, functors	Composition of morphisms, identity, functors, natural transformations	Focus on structure-preserving mappings and how transformations compose, treating relationships between representations as fundamental.
Dynamical	Conceptual states change over time according to systematic transition rules or underlying dynamics.	States, trajectories, transition functions, dynamical systems	State transitions, trajectories, attractors, stability, temporal composition	Understand concepts and representations as processes rather than static objects, emphasizing trajectories, transitions, stability, and change.
Optimization	Behavior or representations can be characterized as solutions to objectives subject to constraints and trade-offs.	Objectives, constraints, feasible sets, decision variables	Minimization/maximization, constraint satisfaction, regularization, gradient-based optimization	Explain conceptual structure in terms of what a system is optimizing and the constraints under which it operates.
Game Theory	Conceptual behavior emerges through strategic interaction among agents with potentially different goals and information.	Agents, strategies, utilities, payoffs, equilibria	Best response, strategic interaction, equilibrium, coordination, incentive optimization	Understand conceptual behavior through incentives and strategic dependencies: what an agent does depends on what other agents do and what outcomes they seek.

Table 4: Conceptual overview of the ten mathematical frameworks considered in our comparison. For each framework, we summarize its core assumption, basic mathematical objects, canonical constructions, and guiding philosophy. These descriptions provide the interpretive assumptions underlying the operational comparison in Table 2; they are intentionally high-level and do not exhaust the mathematical variants or applications of any framework.

Theory	Sup.	Trained	Additional Data	Methodology	ARI	AMI	NMI	Matched Acc.	Acc.
Optimization	○	Yes	Embeddings	Logistic-regression probe (multinomial cross-entropy + L2) by ERM; nested grouped CV	0.717	0.767	0.793	0.833	0.833
Probability	○	Yes	Embeddings	Gaussian class-conditionals with shared covariance (LDA, Ledoit–Wolf shrinkage) on PCA; posterior $P(\text{cat} \mid \text{item})$	0.672	0.726	0.759	0.801	0.798
Geometry	✓	No	Embeddings	k-means (Euclidean, `n_init`=10) at k = number of human categories	0.400	0.591	0.629	0.607	–
Graph	✓	No	KG (WordNet)	k-medoids on WordNet shortest-path distance (max over senses)	0.369	0.528	0.569	0.552	–
Information	○	No	Embeddings and KG (WordNet)	Agglomerative Information Bottleneck: greedy merges minimising loss of $I(T;Y)$	0.363	0.483	0.529	0.538	–
Dynamical	○	No	Embeddings over all checkpoints	k-medoids on rank-normalised cosine distance averaged across 56 checkpoints; settling from nearest-neighbour sets	0.355	0.466	0.514	0.541	–
Logic	○	Yes	Feature norms	Per-category majority-vote necessary/-sufficient feature rules; nested grouped CV	0.335	0.460	0.546	0.560	0.547
Algebra	○	No	Feature norms	Formal Concept Analysis; k-medoids on Galois-closure distance	0.193	0.297	0.419	0.394	–
Category	○	No	KG (WordNet)	Weisfeiler–Lehman refinement (graded bisimulation) on typed WordNet graph; k-medoids on WL-kernel distance	0.062	0.149	0.220	0.249	–
Game-theory	×	–	–	–	–	–	–	–	–

Table 5: Mathematical theories of categorization evaluated against human category judgements. **Sup.**: ✓=supported by the plan, ○=requires some additional data or training, ×=not testable. Scores are means across the three human datasets weighted by item count; item counts differ by theory (952 for embedding-based theories, 938 for WordNet-based, 722 for Logic and 569 for Algebra, reflecting coverage of the required resource). **Matched Acc.** is Hungarian-matched cluster accuracy and is available for all theories; **Acc.** is held-out predictive accuracy and is defined only for the trained theories. The two are not comparable. ARI and AMI are chance-corrected; NMI is not and rises with cluster count. Embeddings were mean pooled extrected from the 24th layer of the final checkpoints using OLMo-7B.

Method	Places *crowbar* with	Representative co-members
Human ground truth: Carpenter's Tool		
Optimization	Bird	crane, robin, sparrow, bluejay
Probability	Bird	robin, sparrow, bluejay, dove
Geometry	Birds	crane, robin, sparrow, canary
Dynamical	Birds	(bird cluster, 48/65)
Logic	Weapon	gun, pistol, revolver, knife, dagger
Graph	Clothing	piano, cushion, cupboard, mirror
Algebra	Toy	table, bench, cushion, fan
Category	—	singleton; no co-members
Information	Carpenter's tool	stereo, television, radio, telephone

Table 6: Where each account places *crowbar*.