FORMAL CONCEPTUAL BLENDING AS A FUNDAMENTAL TOOL FOR ARTIFICIAL MATHEMATICAL INTELLIGENCE AND FOR THE FINE-TUNING OF LARGE LANGUAGE MODELS

DANNY A. J. GÓMEZ-RAMÍREZ AND MARK TURNER

ABSTRACT. We propose, as a foundational pillar of Artificial Mathematical Intelligence, the first global Taxonomy of Seminal Cognitive Operations, that is, operations that structure the mathematical capabilities of the human mind. One of the central cognitive operations of the mathematical mind is general conceptual blending. We provide an elementary overview of how formal models of conceptual blending have been applied to generate and guide reasoning patterns in general mathematics, and how this application of some principles of conceptual blending has led to a quantum jump in the construction of artificial mathematical co-creative agents. Our aim is to fill the long-standing gap in understanding how formal conceptual generation occurs in formal reasoning. This advance also enables us to present simple examples of how blending can improve the training of Large Language Models through fine-tuning: this fine-tuning includes prompting for solutions of mathematical problems ranging from elementary to highly difficult, a chain-of-thought process that relies on instances of formal conceptual blending, and the delivery of a sound final response.

GENERAL INTRODUCTION

Blending is a domain-general cognitive process that takes two or more inputs (e.g. semantic frames, words, constructions, gestures, images, ideas, etc.) and combines them into a blend. This process is selective (not every detail from an input space appears in the blend) and frequently gives rise to emergent meanings (which are not part of any of the input spaces). In the history of philology, rhetoric, and linguistics, there were many efforts to account for what looked like pyrotechnic special cases of blending, such as remarkable metaphors, metonymies, analogies, and so on. But blending is a domain-general process indispensable for all human cognition, including the most basic cases of mathematical invention. This article cannot provide a stand-alone encyclopedic review of Blending Theory (for that, see, e.g. book-length treatments such as Fauconnier & Turner, 2002; Turner 2014). Rather,

Key words and phrases. Formal Conceptual Blending, Artificial Mathematical Intelligence, Cognitive-Computational Metamathematics, Metaphorical Reasoning, Large Language Models.

we present evidence for our claim that formal conceptual blending is a powerful tool for artificial mathematical intelligence.

The work of Conceptual Blending has been extensively examined in art, music, religion, mathematical insight, scientific discovery, advanced social cognition, advanced tool invention, fashion, and so on. Consider the invention of non-Euclidean geometry (Turner 2001), which began in attempts to derive rather than assume the parallel axiom. So many geometers, regarding the parallel axiom as an embarrassment, had tried and failed to prove (rather than assume) the parallel axiom that as early as 1759, d'Alembert referred to it as "the scandal of the elements of geometry." Gerolamo Saccheri (1667-1733) made a remarkable attempt that focused on a quadrilateral ABCD where angle DAB and angle ABC are right angles, and where line segments AD and BC are equal. Without using the parallel axiom, it is easy to prove that angles BCD and CDA must be equal. Saccheri did this. If we assume the parallel axiom, BCD and CDA can be proven to be right angles. (In fact, the implication works in both directions-—the parallel axiom is equivalent to the assumption that BCD and CDA are right angles—but all we need for this analysis is the first direction: the parallel axiom implies that BCD and CDA are right angles.) Therefore, if we deny that BCD and CDA are right angles, we thereby deny the parallel axiom. Saccheri did just this, in the hope of deriving a contradiction from the denial, which would prove the parallel axiom by *reductio ad absurdum*. (See Figure 1).

But if BCD and CDA are not right, they are still equal, and so they must be either obtuse or acute. Saccheri attempted to show that, in either case, a contradiction follows. He assumed that they are acute; that is, he proposed a conceptual blend of two routine Euclidean figures. Both have a quadrilateral ABCD, equal line segments AD and BC, equal angles DAB and ABC, and equal angles BCD and CDA. The blend takes this structure from both inputs. But the first input has right internal angles DAB and ABC, and the second input has acute internal angles BCD and CDA. The blend takes the right angles from the first input and the acute angles from the second.

The blend is impossible in Euclidean geometry, but Saccheri never found a contradiction for it. In fact, he produced many astonishing theorems that are now recognized as belonging to hyperbolic geometry. These theorems were so repugnant to commonsense notions that he concluded that they must be rejected. It is important to see that all of Saccheri's elaboration of the blend followed everyday procedures of Euclidean geometry. The input spaces are Euclidean and familiar; the elaboration procedures are Euclidean and familiar. The only new thing in the process is the selective, two-sided projection to create the blend, and the emergent structure in the blend. One of the theorems Saccheri derived was another blend having to do with parallel lines: given any point A and a line b, there exist in the pencil (family) of lines through A two lines p and q that divide the pencil into two parts. The first of

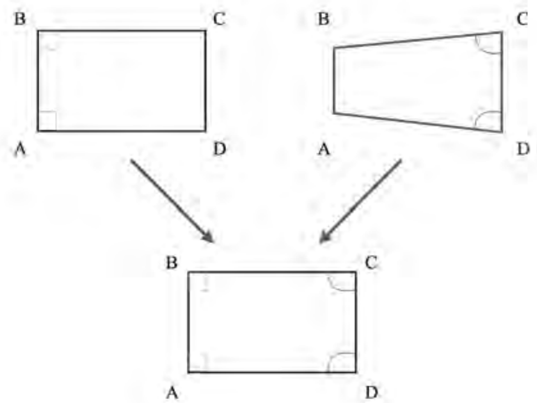

FIGURE 1. Non-Euclidean Quadrilateral

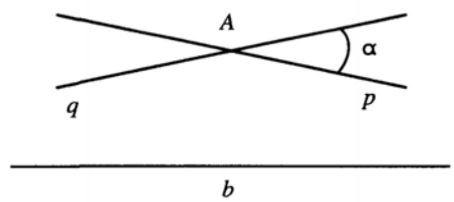

FIGURE 2. Pencil of Lines

these two parts consists of the lines that intersect b, and the second consists of those lines (lying in angle α) that have a common perpendicular with b somewhere along b. The lines p and q themselves are asymptotic to b. In Euclidean geometry, angle α is 0 and p and q are identical. But for Saccheri's new geometry, derived from the acute-angle blend, angle α is positive and p and q are distinct, so there is an infinity of lines through A that have a common perpendicular with b and never meet it. This is a blend of two standard notions: the first is a schema of a parallel line through a point outside a line; the second is a schema of a bundle of lines through a point. If we blend these, we have multiple lines through a point outside a line that are all parallel to the line.

Saccheri is not credited with the invention of non-Euclidean geometry. Credit is given instead to Gauss, Bolyai, and Lobatchevsky for recognizing (but not proving) that hyperbolic non-Euclidean geometry is mathematically consistent, and to Gauss for recognizing that physical space might be non-Euclidean. For us, today, non-Euclidean geometry is indispensable to science. We must use it, for example, to compute how to land a craft on Mars. The theory of general relativity is fundamentally built upon non-Euclidean geometry. Non-Euclidean geometry, particularly Riemannian geometry, which generalizes to curved spaces, provides the necessary mathematical tools to describe the motion of planets around stars, the bending of light around massive objects (gravitational lensing), and the behavior of objects near black holes. The operation of conceptual blending in mathematics is pervasive. It is extensively studied in (Alexander, 2011; Lakoff & Núñez, 2000; Núñez in press; Steen & Turner 2025, Turner 2019).

Blending is often introduced through such highly visible, pyrotechnic examples, for the sake of pedagogy. But using flashy examples is deeply misleading. Nearly all conceptual blending goes completely unnoticed in consciousness. By contrast, *the cyclic day* provides an example where the blending is still visible, but, typically, only once it has been pointed out. An actual day is an experience, and we can think of days as events along a time line. That is already impressive blending: days do not lie upon a line; nor do we move along a line through them (see Fauconnier & Turner 2008 for an analysis of the role of blending in conceiving of time). Further blending compresses a sequence of individual days into *the cyclic day*. The analogies across the inputs are compressed to identity (*the* cyclic day) and the differences are compressed to change. Now, in the blend, but not in the inputs, there is the amazing emergent structure that the day *repeats*. This conceptual blend simplifies navigation through daily routines, cultural practices, and societal structures.

The site at `http://blending.stanford.edu/` cites many hundreds of such studies of conceptual blending, in cognitive science, linguistics, psychology, neuroscience, psychotherapy, physics, political science, education, film, pretend play, religion, artificial intelligence, computation, modeling, data science, gesture, music, painting, comics, counterfactual thinking, reasoning, logic, inference, and many other fields outside of mathematics.

There have been studies of the extent to which the mental operation of blending is available to other species, such as the special panel on "Conceptual Blending in Animal Cognition" at the conference of the Cognitive Science Society, in Vienna, 26-29 July 2021 `http://youtu.be/aPpmCX0Tl7w`. Many operations of integration have been studied in neuroscience (e.g., sensory, perceptual, time-space collocation, object permanence).

Cognitive science analyzes cognitive operations, but it as not as if we *a priori* have an inventory and analysis of them. What are the operations of the human mind? To be sure, we have mere *words*, such as "attention," "reason," "memory," and so on, but they are only a shorthand to help us get rolling. It takes theoretical research to establish what cognitive operations exist and how they work. There is no prior established definition of "blending" that we can take for granted. Evidence of creative conceptual combination and elaboration prompts us to seek the underlying cognitive operations. This research program is referred to as "Conceptual Integration Theory," also known as "Blending Theory." Cognitive science begins with exploratory concepts of mental operations and revises them as the science progresses.

There have been various attempts to provide mathematical models of mind. It is standard for a science to use mathematics as a platform for modeling whatever phenomena it studies. Mathematics is, for example, used to give a model of Newtonian mechanics; rather different mathematics is used to give a model of quantum mechanics. Our purpose in this article has nothing to do with giving a mathematical model of blending. Indeed, it is unlikely that mathematics yet possesses the tools that would be needed to model such a complicated, flexible, and powerful system, one that operates in both individual agents and, in distributed fashion, across networks of agents. Our purpose runs in the opposite direction: we study how to improve mathematics by adding to mathematics some of the insights from Blending Theory. In what follows, we will explore a formal metamathematical system, one with strict definitions. But this system is not a mathematical model or representation of the operation of blending, or a substitute for or refinement of theory of blending. It is not a refinement of theory of blending. Instead, we explore how cognitive-computational metamathematics might be advanced by borrowing from theory of blending.

We move now to a discussion of formal conceptual blending in cognitively-inspired artificial mathematical invention. Notation for blending diagrams varies. Most research on blending offers pedagogical diagrams in which the blended mental space is placed at the bottom of the diagram. (In hyper-blending, where a blended space is itself an input to a further blend, diagrams can get complicated.) In contrast, in the mathematical field of abstract category theory, the graphical representation of a "pushout"–that is, a kind of gluing together of two objects along a shared sub-part–is placed at the *top* of the graph. In category theory, a pushout is a specific type of colimit. It starts with a setup of three objects (X, Y, and Z) and two maps (morphisms) going out from Z:$f : Z \to X$ and $g : Z \to Y$. The pushout is a new object, let's call it P, along with two new maps ($X \to P$ and $Y \to P$) that complete a "commutative square." The pushout represents the most efficient, universal way to combine X and Y based on their relationship to Z, without adding any extra unnecessary information. This category theory concept of a pushout matches only slightly the cognitive scientific concept of a blend. Blending, as a mental operation, is not algorithmic and can involve many input spaces. The same set of inputs can result in different blends, depending on the selective projection and the focus of emergence. But in what follows, we will switch to the category theory convention of placing the blend-pushout at the *top* of the diagram rather than the *bottom*. We generalize this formalitazion further to the meta-logical level in the setting of the category of formal concepts.

1. Foundational Properties of Conceptual Blending

In this section, we present informally and generally principles that we take up technically in section 6.1 of the paper.

Human thought is highly conservative, rooted in concepts we already know, but it is also remarkably creative. Blending theory proposes that we do not merely abstract or analogize across concepts but instead *blend* what we already know to create *new emergent structure in the blend* that is useful because it is *tractable* mentally and can be carried within our mental competence and *expanded* as needed to think about particular situations. Suppose we are actually thinking of the real world. In particular, we are thinking of *my brother-in-law, the stockbroker*. Following Fauconnier, we would say that we are activating a *mental space*.

Mental space: a mental space is a small array of related mental elements that one can activate simultaneously in the mind. In this case, ”My brother-in-law, the stockbroker” prompts us to activate a mental space with the elements *one man (I, who is the speaker)*, *another man A*, the role *stock-broker*, a role-value connection from the role *stock-broker* to the value *A*, and a *relation of brother-in-law* between *I* and *A*. In cognitive science, such a mental space is usually represented as a circle with elements and relations within it. Mental spaces exist in mental webs.

Mental web (alternatively called a mental network, or conceptual network; many other metaphors are available, each with its own deficits): A mental web is a set of mental spaces that are activated and connected as one is thinking about a topic. For example, ”My brother-in-law, the stockbroker, will be traveling from San Francisco to Cleveland for Thanksgiving for a massive family reunion, and we need to learn the time of his arrival so that we can go pick him up” will prompt for many mental spaces, such as a mental space in which I locate my car keys, another in which I travel to my car, another in which I drive through no doubt complicated holiday traffic, another in which I stop at the arrival deck of Cleveland Hopkins airport at gate 5, and on and on. Typically, one cannot hold all these spaces equally active simultaneously in the mind. We will focus on one or another mental space at a time, but activate many as we think, and those we have activated recently will remain latent, which is to say, easier to activate.

Vital Relations: The mental web will have many conceptual connections between the individual mental spaces in the web. The most frequent and important mental connections are the ”Vital Relations”: *Time, Space, Identity, Change, Cause-Effect, Part-Whole, Representation, Analogy, Disanalogy, Representation, Property, Similarity, Category, Intentionality,* and *Uniqueness*. For example, in the mental web about *my picking up my brother-in-law and family* at the airport for the massive family reunion, there will be an agent in several of those mental spaces corresponding to *I*, and all of those elements in all of those mental spaces will be connected by *Identity* connectors. These mental spaces in the mental web will be connected by many other connectors. The pickup at the airport is connected by a *Time* connector so that the pickup is suitably prior in time to the mental space in which we all have the Thanksgiving feast. And that pickup at the airport is connected by a *Space* connector so that we understand that the airport is at a spatial remove from the house in which the Thanksgiving feast will be held in a vast the vast dining room.

But now comes the crucial step for new concepts. We can *blend* mental spaces in this mental web.

Blend. A *blend* is a mental space that results from *blending* mental spaces in this mental web. The blend is not an abstraction, or an analogy, or anything else already named and recognized in common sense, although blending is the basis of the cognitively modern human mind. A *blend* is a new mental space that consists of a partial combination of elements from different mental spaces but that develops new conceptual structure of its own not drawn from those spaces. For example, ”If I were a stockbroker, like my brother-in-law, who lives in San Francisco and so must rise every day the market is open at 5am, I would be miserable” asks us to make a mental space in which there is one man (*I*) who is imbued with some of what we think about the speaker and some of what we think about the brother-in-law, but only some in each case, and for whom we develop new structure. In the blend, there is an element that has the personal identity of the speaker, but no longer has that person’s job. This element is not available from any other space in the mental web. It is unique to the blend. This element is emergent in the blend. There are many other elements and properties in the blend that are not available from other spaces in the mental web (or network, or whatever metaphor one prefers for a set of connected ideas). Neither the speaker nor the brother-in-law is miserable in the mental spaces that are blended, but in the blend, the person is miserable. This is new in the blend. The mental spaces in the mental web that contribute to the creation of the blend are called by a number of names: ”inputs,” ”contributors,” ”donors,” etc. All of these metaphoric names have their own communicative deficits.

Projection. The elements and relations that come into the blend from the mental spaces that are blended are referred to as *projections*. These projections to a blend are always *partial* or rather *selective*. For example, in ”If I were a stockbroker, like my brother-in-law, who lives in San Francisco and so must rise every day the market is open at 5am, I would be miserable,” we project to the blend the speaker but only a small part of what we know about the speaker. We do not project the speaker’s current employment, for example, because then the speaker could not be a stockbroker. We project from the mental space with the stockbroker brother-in-law the role *stockbroker* and perhaps even *living in San Francisco and accordingly rising every weekday at 5am*, but not of course the physical appearance of the brother-in-law, or his family relations, and so on. We might project his lodging, but we might not, and this variation illustrates the fact that what we project to the blend as we build it is largely left up to the different members of the communicative scene. Disagreements and misunderstandings naturally arise. Blends are sometimes hotly contested in the culture, as we see in the current turmoil over how to conceive of the blend *intellectual property*.

Emergent structure in the blend and in the mental web: In the blend, the speaker is a stockbroker and is miserable. In no other space is any of this true. This structure is *emergent* in the blend. It arises as a consequence of making the blend. (There are many kinds of emergent structure discussed in (Fauconnier Turner 2002) and elsewhere, e.g. emergent structure from "composition," "completion," and "running the blend.") Crucially, there is also new emergent structure in the web outside of the blend. Now we know, for example, that the speaker in his own reality has an aversion to rising early. This is new structure we build for the speaker in his own reality that we did not have there before. We build it for the speaker as a consequence of what we have learned by building the blend. There is also emergent structure in the connection between the speaker in his input mental space and the stockbroker in his input mental space, namely a *disanalogy* connection between them. Of course, there always were disanalogies between them in the mental web: the brother-in-law lived in San Francisco but the speaker presumably lived somewhere else, like Cleveland; the brother-in-law was a stockbroker but presumably the speaker was not; and so on. But now there is a *new, emergent* disanalogy connection between these spaces, having to do with disposition. It could have been different, and would have been if the sentence had been "If I were a stockbroker, like my brother-in-law, who lives in San Francisco and so must rise every day the market is open at 5am, I would be happy." When we project what we have learned in the blend back to original mental spaces in the mental web, we do not always project exactly what has arisen in the blend.

Human-scale: Some bundles of thought are tractable and managable by the human mind. We call them *human-scale*. Other bundles of thought are not tractable, because we cannot grasp them mentally, or they go beyond our mental limits. Most mental webs for what we must think through would be utterly intractable for us were it not that we can make a human-scale blend drawing on different mental spaces in the web. The blend then gives us a tractable thing to think about. It helps us access, organize, manipulate, and adjust the mental web in which it now sits. For example, in the vast mental web of thinking about life and possibilities, one can have a little blend in which one actually is a stockbroker, going through the motions. The blend in this case is a kind of human-scale mental simulation.

Compression: A blend is not a small abstraction of the mental spaces it blends and is not a partial cut-and-paste assembly, either, because it contains emergent structure. It is a packed little *compression* containing much less information than is contained in the mental web it serves, and in particular a compression from which we can access the rest of the mental web in which it sits.

Expansion: We use compressed, tractable blends to help us think about larger mental webs, or mental networks. We might say, metaphorically, that we carry small, compressed blends with us mentally, and *unpack* or *expand* them as needed to connect up to what we need to think about. For example, the pithy, compressed little blend with the miserable stockbroker can be used to help the speaker think about any job in a time zone other than the Eastern time zone (GMT - 5) and help him be vigilant about inquiring into demands that any job might impose because of events that lie outside its time zone.

This example about the stockbroker is provided to furnish some vocabulary, but is is crucial not to over-generalize from it. Although this example is obvious as a blend, essentially all blending is invisible to consciousness and does not look at first like blending. The counterfactual ”If I were . . .” is only one very tiny example of a linguistic prompt for blending, and many other kinds of products arise from the operations of blending. Blends need not be counterfactual, hypothetical, false, or fictional. On the contrary, many blends constitute what we feel in common sense to be ground truth, as we saw in the example of the *cyclic day*, which arises by using a particular generic blending template, one that is very common: There are analogies and disanalogies across the different days in our experience. The analogies are packed to one thing in the blend: the *day*. The disanalogies are packed to change for that thing: the day is *cyclic*; it *starts over* every dawn and *repeats*. It is crucial to recognize that the cyclic day blend is not just an abstraction. There is emergent structure in the blend that is not in any of the individual input days, and accordingly it is not an abstraction. Indeed, almost no blends consist just of structure that is equally shared by the inputs to the blend. For example, no one of the individual input days to the cyclic day blend is cyclic; no one of those individual input days repeats or starts over. But because of blending, all the days that have ever happened or will happen can be packed into a single conceptual unit—the *cyclic day*, which *repeats*. Thinking of the cyclic day, we can say, “dawn is coming around *again*,” or “It’s time for my *morning* coffee” or “this park closes *at dusk*.” There can be morning coffee even in a day in which we did not have any coffee. We know how these words and concepts apply to the compact blend, and we can unpack or expand from that blend to any parts of the network that interest us, to any day at any time, and even to a

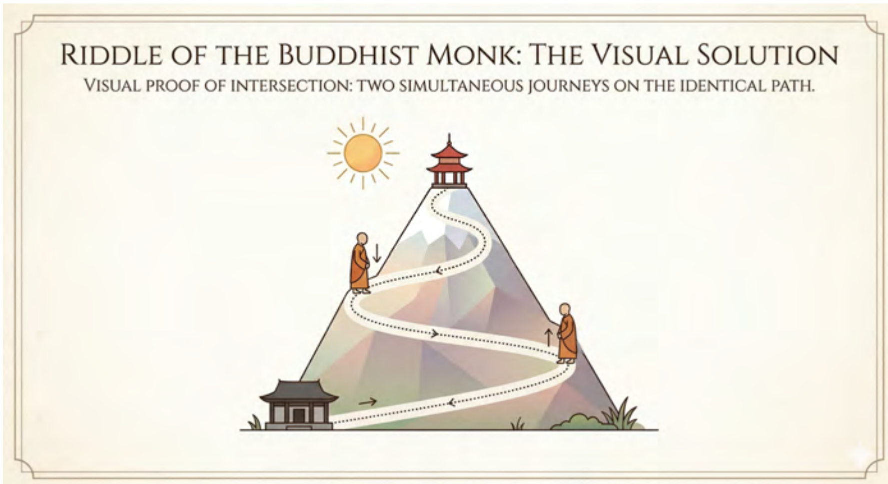

FIGURE 3. Riddle of the Buddhist Monk

tractable stretch of days. The cyclic day is a compact blend, a congenial mental space for thinking about the vast sequence of days. This vast sequence of days is of course itself too big for working memory. The compact blend—the single *cyclic day*—makes it possible for us to work with concepts of time that stretch far beyond what we would otherwise be able to manage.

We can give a quick diagram illustrating all these ideas—mental space, mental web, connectors between spaces, emergent structure, projection, compression, expansion, human-scale, blend. It is a diagram from Fauconnier & Turner (2002). Consider the riddle of the Buddhist Monk:

> A Buddhist monk at the foot of a mountain path a little before dawn thinks of climbing the mountain. At dawn, he begins walking up the path. He reaches the top at sunset, meditates at the top overnight until, at dawn, he begins to walk back to the foot of the mountain, which he reaches at sunset. Make no assumptions about his starting or stopping during the scent and descent or about his pace during the trips. Riddle: is there a place on the path that the monk occupies at the same hour of the day on the two trips?

One way to solve this riddle is to blend the monk's ascent with the monk's descent, so that in the blend, at dawn, there are two monks, one at the foot of the mountain, the other at the top. They then take their journeys, each arriving at the opposite end of the path at sunset. They must meet somewhere, and where they meet is the spot on the path that they occupy at the same hour of the day on the two separate journeys.

The connected set of ideas for solving this riddle is a *mental web*. It contains at least three *mental spaces*. There are connectors between mental spaces, such as identity connectors between the path in the mental space for the ascent and the path in the mental space for the descent. Some but not all the information from those two mental spaces is *projected* to a *blended* mental space. We do not, for example, project the date of the ascent and the date of the descent, or the weather on those days, or the fact that the monk is aware of what is around him and would surely be shocked to find himself approaching himself on the path. We do not project the fact that a person cannot be in two places (foot and summit) at the same time. The blend is a *compression* of parts of its mental web, and it is at *human-scale* because it is a little vignette about two people approaching each other on a path; this is a simple and familiar scene of human walking. But this compressed blend also has *emergent* structure. It has two monks, and a meeting. We can use the compressed blend to think about and work on the mental web. We can *expand* or *unpack* the blend and connect it back up to elements in the larger mental web. Some of the emergent structure in the blend, namely, the fact that there is a meeting, leads us to project back to create new structure in the mental web itself: now, for example, there is an identity connection between some spot on the path in the ascent mental space and a spot on the path in the descent mental space such that the monk is located at that spot in his travel at the same time of day on the two separate days. The integration network might also contain a generic mental space, which is dynamic, and which captures at any moment some structure shared across the input spaces. The generic mental space changes as inputs are reconsidered and modified. In the case of the Riddle of the Buddhist Monk, the generic space for the network that solves the problem has a monk traversing a mountain path from one end to the other in the interval from dawn to sunset, but not the direction, position, or specific day.

Figure: Several different types of conceptual integration networks were given particular Greek or Latin names. For a full analysis, see (Turner 1998, ”Figure,” chapter 2 of Cristina Cacciari, Ray Gibbs, Jr., Albert Katz, and Mark Turner. *Figurative Language and Thought*. New York: Oxford University Press, 1998). People do tend to feel that there is a difference between metaphorical and non-metaphorical thought and expression, and it is important to account for that feeling. What people recognize as metaphor are blends with some typical features: (1) strong asymmetry. One input (the “source”) contributes more concrete or imageable structure; the other (the “target”) is more abstract. (2) For cliché metaphors, historically exported vocabularly from the ”source” input, so that it is common to speak of the blend in robust vocabulary native to the source domain. In the blending analysis, metaphor is ubiquitous but not special; creativity is central, not peripheral; meaning is constructed by blending, not transferred; understanding a metaphor is building a blend, not decoding a mapping; metaphor is not a primitive operation but a highly typical, asymmetrical, and compressive form of conceptual blending that yields emergent meaning through selective projection across mental spaces.

2. Seminal Pillars of Cognitive-Computational Metamathematics or Artificial Mathematical Intelligence

Artificial Mathematical Intelligence (AMI), also referred to as Cognitive-Computational Metamathematics (CCMM) [8] is an ambitious research program for integrating formal conceptual blending as a seminal co-creative mechanism. It seeks to develop a Universal Mathematical Artificial Agent (UMAA) capable of addressing a substantial portion of open mathematical problems. The goal is for the agent to be able to resolve —either by producing a rigorous formal proof or by constructing a valid counterexample—approximately 80% of the conjectures, partially solved questions, and symbolic mathematical challenges presented to it.

Within the AMI framework, this goal is pursued through a bottom-up methodological strategy. The first foundational component of this approach is the research program known as the New Cognitive Foundations of Mathematics [10]. The central objective of this program is to formulate deeper and more structurally robust refinements of the fundamental conceptual frameworks that underpin contemporary logic and mathematics. For instance, improved formal characterizations must be developed for notions such as first-order and higher-order logic, sets and membership relations, various number systems (natural, integer, rational, and real numbers), as well as mathematical objects such as continuous functions or algebraic groups. The ultimate aim is to enable a precise computational encoding of the syntactic and semantic structures underlying these concepts, thereby allowing mathematical conjectures to be systematically translated into an appropriate computational language and implemented as executable procedures.

The second foundational pillar of AMI consists of the systematic formalization and refinement of a universal taxonomy of cognitive-computational metamathematical mechanisms employed by mathematicians during advanced research. An initial large-scale taxonomy comprising roughly twenty core mechanisms is proposed in [8, Part II]. This metacognitive framework is essential because AMI adopts a "thinking humanly" paradigm in artificial intelligence [3]. In practical terms, the CCMM program attempts to reproduce the types of mental operations that mathematicians employ when generating and manipulating concepts—both established and novel—through processes such as exploratory sketches, illustrative examples, trial-and-error experimentation, metaphorical reasoning, analogical transfer, conceptual blending, and selective specialization [12], [13]. [1]

A third structural component of the AMI program involves the development and integration of the computational infrastructure necessary to construct preliminary instances of a UMAA within specific mathematical domains, including elementary number theory, abstract algebra, and real analysis [11], [13], [16]. In this context, the notion of a pseudo-pre-code, also described as a cognitive-computational proof, emerges as a richer and more multidimensional extension of the traditional mathematical proof. As an initial validation of the previously mentioned taxonomy of cognitive mechanisms, [13] provides explicit pseudo-pre-code constructions for more than thirty-four mathematical notions and theorems. These examples range from elementary objects such as the concept of a mathematical function to sophisticated structures like the notion of a scheme in modern algebraic geometry [17, Ch.2].

A particularly distinctive contribution of the AMI (CCMM) framework lies in its attempt to address the long-standing gap concerning the artificial generation of abstract mathematical concepts inspired by human cognition [16]. To date, neither logic nor computer science has produced a comprehensive causal framework explaining how advanced mathematical concepts, structures, and theoretical systems arise within the human mind. Although earlier proposals—such as the influential work of Lakoff and Nu nez based on metaphorical cognition and analogical reasoning [18]—offer important insights from a cognitive perspective, their explanatory scope remains relatively limited from a pragmatic and global standpoint. In actual mathematical practice, researchers rely on a much richer repertoire of cognitive heuristics. These include the systematic use of examples, processes of generalization and specialization, conceptual identification and duplication, and other forms of conceptual transformation, all of which play a crucial role in everyday mathematical creativity [12].

Another significant strength of the AMI paradigm is its inherent compatibility with existing formal-verification technologies. Rather than competing with current tools, CCMM is intended to complement and enrich a wide range of systems, including formal verification languages, interactive proof assistants, automated theorem provers, SAT and SMT solvers, computer algebra systems, and related computational platforms. In a broader perspective, AMI proposes the eventual development of a unified computational framework—potentially involving a sufficiently expressive mathematical language and an interactive software environment—capable of addressing complex research problems and presenting solutions in a conceptually transparent and human-interpretable manner.

[1] It is worth noting that such an approach differs substantially from the internal design principles of current large-scale language models such as ChatGPT, DeepSeek, or Gemini.

Accordingly, the purpose of the present work is not to evaluate the relative strengths or weaknesses of existing computational tools used in formal mathematical research. Rather, our objective is to explore how certain heuristic and formal principles emerging from the AMI framework may contribute to the refinement and enhancement of such systems, particularly in the design and organization of specialized benchmark datasets for advanced pure mathematics. The long-standing vision articulated by Alan Mathison Turing—the creation of an artificial mathematician capable of collaborating with humans in mathematical discovery [24], [15]—will likely require not only technological competition but also sustained cooperation across the academic, scientific, and entrepreneurial communities.

For readers interested in the broader landscape of automated reasoning systems and artificial intelligence applied to formalizable mathematics, several comprehensive surveys provide valuable context. Notable examples include [1], [19], [23], [21], [2], and [4], which collectively review major developments in formal verification, theorem proving, and computational mathematics.

Finally, a further advantage of the AMI paradigm lies in its potential to serve as a conceptual integrative layer across diverse formal systems. In principle, CCMM can operate alongside—and potentially unify—the capabilities of existing proof assistants, verification languages, automated reasoning engines, and symbolic computation systems. The long-term ambition is to develop a sufficiently expressive computational environment capable not only of solving complex mathematical problems but also of communicating those solutions in a manner that remains both mathematically rigorous and cognitively transparent.

3. Metamathematical Preliminaries

The purpose of this section is to establish the minimal metamathematical vocabulary required for the formal treatment of metaphor, conceptual blending, and cognitively structured mathematical reasoning developed below. The objective is not to propose a new foundational logic, nor to reduce mathematical creativity to a fixed calculus. Rather, we introduce a sufficiently general descriptive layer in which mathematical objects, their explicit presentations, the operations performed upon them, and the trajectories generated by those operations can be distinguished without being artificially separated.

This distinction is indispensable for the present work. A mathematical proof may be correct at the object level while its discovery depends on transformations that are not themselves steps of the proof: choosing an example, suppressing a hypothesis, changing notation, transporting a relation to another domain, identifying a common pattern, or constructing an auxiliary concept. Such actions are *metamathematical* in the operational sense adopted here. They act on mathematical descriptions, concepts, examples, conjectures, and proof states, and they modify the conceptual conditions under which subsequent deductions become available. The same distinction is crucial computationally: a system may manipulate formulas correctly while failing to preserve the objects, dependencies, and conceptual commitments that give those formulas mathematical meaning.

3.1. **Object-level and metamathematical levels.** Let L be a logical environment sufficiently expressive to support appropriate notions of signature, well-formed expression, inference, satisfaction, and model. The framework is deliberately logic-neutral: depending on the mathematical domain, L may be many-sorted first-order logic, a higher-order system, a type theory, an institutional semantics, or a less fully formal but mathematically controlled representational language [6]. What matters here is not the selection of a privileged foundational system, but the possibility of distinguishing meaningful mathematical expressions from the transformations performed upon their organization.

We shall use the term *object-level activity* for reasoning carried out within a fixed mathematical environment. Examples include deriving a consequence from a collection of hypotheses, evaluating a term, verifying that a structure satisfies an axiom, or constructing an object whose required type and defining conditions have already been specified.

By contrast, a *metamathematical operation* changes, selects, compares, reorganizes, or generates some component of the environment in which object-level reasoning occurs. Typical instances include:

(1) selecting a concrete or generic example of a concept;
(2) strengthening or weakening a hypothesis set;
(3) replacing one symbolic or conceptual component by another;
(4) identifying a shared relational pattern across different domains;
(5) generalizing, particularizing, or duplicating a mathematical structure;
(6) projecting selected information from several conceptual inputs into a new integrated structure;
(7) changing the active representation of a problem while preserving all information relevant to its solution.

The boundary between the two levels is functional rather than absolute. An operation that is metamathematical relative to one task can become an explicitly formal object-level construction in a richer theory. Accordingly, the distinction should always be read relative to a working mathematical environment and to a current cognitive goal. This relativity prevents a common category error: treating the discovery process as if it were merely an incomplete proof, or treating every valid proof transformation as if it explained how the relevant proof architecture was generated.

Definition 3.1. A *mathematical working unit* is any cognitively addressable entity that can occur as an input to, or an output of, a mathematical reasoning operation. It may be a symbol, term, formula, diagram, example, model, definition, conjecture, proof fragment, proof strategy, mathematical concept, or a structured collection of such entities.

The expression "cognitively addressable" means that the unit can be selected and manipulated as a sufficiently coherent item within the reasoning process. It does not imply that the unit is atomic, consciously represented in full detail, or encoded in a unique format. A proof strategy, for example, may be treated as one working unit at a global level while containing many subordinate units when it is expanded.

3.2. **Presentation, semantic content, and conceptual organization.** Mathematical cognition does not operate on semantic content in isolation. It also depends on the particular symbolic, diagrammatic, linguistic, and spatial forms through which that content is made available. Two extensionally equivalent formulations can differ substantially in their heuristic value: one may expose an invariant, suggest a useful analogy, or make a decomposition visible, while the other conceals the same structure. Consequently, the framework must retain both semantic and representational information.

Definition 3.2. A *mathematical description* is a triple

$$\mathfrak{D} = \langle \mathsf{Pres}(\mathfrak{D}), \mathsf{Cont}(\mathfrak{D}), \mathsf{Role}(\mathfrak{D}) \rangle,$$

where:

(1) $\mathsf{Pres}(\mathfrak{D})$ is its explicit presentation, including the relevant symbols, syntactic configurations, diagrams, labels, or linguistic expressions;
(2) $\mathsf{Cont}(\mathfrak{D})$ is its intended mathematical content, determined by the interpretations, models, or inferential commitments associated with the presentation;
(3) $\mathsf{Role}(\mathfrak{D})$ records its current function in the reasoning process, such as hypothesis, target, example, invariant, auxiliary construction, obstruction, or candidate generalization.

The third component is goal-sensitive. The same equation can function as a defining condition in one context, as a derived invariant in another, and as a counterexample-generating constraint in a third. Thus, semantic equivalence alone does not determine cognitive equivalence. This observation will later be essential for conceptual blending: selective projection depends not only on what an input means, but also on which components are active and what function they perform in the integration network.

We do not assume a strict one-to-one correspondence between presentations and contents. A single mathematical content may admit several presentations, while a single presentation may support different interpretations in different environments. We write

$$\mathfrak{D} \equiv_{\mathrm{sem}} \mathfrak{D}'$$

when two descriptions are semantically equivalent for the purposes of the current task, and

$$\mathfrak{D} \equiv_{\mathrm{cog}} \mathfrak{D}'$$

when, in addition, replacing one by the other preserves the task-relevant conceptual roles, inferential affordances, and active dependencies. In general,

$$\mathfrak{D} \equiv_{\mathrm{cog}} \mathfrak{D}' \quad \Longrightarrow \quad \mathfrak{D} \equiv_{\mathrm{sem}} \mathfrak{D}',$$

whereas the converse need not hold.

This distinction provides a formal place for a familiar phenomenon in mathematical practice. Rewriting an expression may preserve truth while changing what can be perceived or inferred from it. Factoring a polynomial, coordinatizing a geometric configuration, or viewing a recurrence as a linear operator can expose new relations without altering the underlying mathematical object. Such changes are not merely stylistic; they can reorganize the available problem-solving space.

3.3. **Conceptual states and cognitive focus.** At any stage of a mathematical task, only a small portion of the potentially relevant information is actively used. A workable metamathematical model should therefore represent not only accumulated content, but also selection, focus, and goal structure.

Definition 3.3. A *conceptual state* is a quadruple

$$Q = \langle \mathcal{W}, \mathcal{F}, \mathcal{G}, \mathcal{R} \rangle,$$

where:

(1) $\mathcal{W}$ is a finite or effectively accessible workspace of mathematical working units;
(2) $\mathcal{F} \subseteq \mathcal{W}$ is the currently focused subcollection;
(3) $\mathcal{G}$ is the active mathematical goal or hierarchically organized family of goals;
(4) $\mathcal{R}$ is a structured collection of task-relevant relations among units in $\mathcal{W}$, including inferential dependencies, analogical correspondences, type constraints, identity links, incompatibilities, and part–whole relations.

The finiteness or effective accessibility requirement is pragmatic rather than ontological. A mathematical theory can contain indefinitely many consequences, but a human or artificial reasoner acts at each stage upon a bounded representation of them. This local character of reasoning explains why selection and compression are not peripheral conveniences: they are constitutive conditions for managing sophisticated mathematical structures.

The focus $\mathcal{F}$ need not contain all the information that is causally relevant to the next step. Some background structure may remain available without being explicitly foregrounded. Conversely, a unit may become focused and later be suppressed without being deleted from the workspace. This difference between *removal* and *deactivation* is important when describing conceptual weakening, strategic backtracking, and the recovery of previously abandoned proof paths.

3.4. **Metamathematical operations and trajectories.**

Definition 3.4. A *metamathematical operation* is a partial transformation

$$\Omega : \mathcal{Q} \rightharpoonup \mathcal{Q},$$

on a class $\mathcal{Q}$ of conceptual states. Its domain consists of the states satisfying the representational, logical, and goal-relative preconditions required for the operation to be meaningfully applied.

Partiality is essential. Generalization may be ill-defined when different occurrences of a component are not sufficiently independent; a metaphorical projection may fail because it violates type or arity constraints; a weakening may destroy a premise required by the target; and a proposed blend may possess no coherent semantic realization. Treating cognitive operations as total functions would conceal precisely the failures that a cognitively grounded benchmark or fine-tuning objective should detect.

An operation may affect one or several components of a state. It can modify the workspace by generating a new unit, redirect the focus, refine the goal hierarchy, or reorganize the relation structure. The elementary operations discussed in the broader AMI taxonomy—including exemplification, strengthening, weakening, comparison, replacement, identification, duplication, generalization, particularization, metaphorical reasoning, and conceptual blending—can all be interpreted as structured families of such partial transformations [12].

Definition 3.5. Given an initial state Q_0, a *metamathematical trajectory* of length n is a sequence

$$\tau = (Q_0 \xrightarrow{\Omega_0} Q_1 \xrightarrow{\Omega_1} \cdots \xrightarrow{\Omega_{n-1}} Q_n),$$

such that Q_i belongs to the domain of Ω_i and $\Omega_i(Q_i) = Q_{i+1}$ for every $0 \leqslant i < n$.

A trajectory is not identified with a formal proof. It may contain exploratory branches, failed transformations, examples used only diagnostically, provisional identifications, or intermediate objects absent from the final exposition. A proof is typically one mathematically certified projection of a richer trajectory. Conversely, the same final proof may be reached through substantially different trajectories. This many-to-one relation between discovery and justification is one reason why final-answer correctness alone gives an incomplete account of mathematical competence.

3.5. **Admissibility and preservation.** Not every formally writable transformation constitutes an acceptable metamathematical step. We therefore introduce conditions that separate a meaningful conceptual transformation from an unconstrained change of notation or content.

Definition 3.6. Let $Q \xrightarrow{\Omega} Q'$ be a transition. Relative to a goal $\mathcal{G}$, the transition is *admissible* when it satisfies the following conditions to the degree required by the operation:

(1) *well-formedness*: every generated presentation is meaningful in its declared representational environment;
(2) *type and arity coherence*: symbolic and conceptual components are combined only in structurally compatible positions;
(3) *semantic viability*: the resulting commitments possess an intended interpretation or an explicitly marked hypothetical status;
(4) *dependency preservation*: any retained conclusion continues to be supported by premises that remain active or accessible;
(5) *goal relevance*: the transformation contributes to, refines, tests, decomposes, or legitimately revises the active goal;
(6) *provenance*: the origin of newly introduced structure can be traced to projection, inference, construction, abstraction, or another declared generative operation.

These requirements are deliberately operation-sensitive. Semantic equivalence is appropriate for a change of representation, but it would be too restrictive for conceptual blending, whose defining purpose is to produce emergent structure. Similarly, preservation of every hypothesis would defeat conceptual weakening. What must be preserved is not an undifferentiated body of information, but the invariant profile appropriate to the transformation.

Definition 3.7. An *invariant profile* for an operation Ω at a state Q is a collection $\mathcal{I}_{\Omega,Q}$ of syntactic, semantic, relational, or strategic constraints that must remain valid under the transition $Q \xrightarrow{\Omega} Q'$. We call the transition *locally coherent* when every element of $\mathcal{I}_{\Omega,Q}$ is preserved in Q' or is explicitly discharged, revised, or suspended by the operation.

This formulation captures a central principle of cognitively structured mathematical reasoning: novelty is compatible with rigor only when change is controlled. The relevant invariant profile may include a type assignment, a commutative relation, an algebraic identity, a geometric incidence condition, the scope of a quantifier, the domain of a lemma, or the global strategy currently governing the proof. A reasoning failure occurs not only when a final statement is false, but also when such a constraint is silently lost during a transition.

3.6. **Traces, explanations, and computational observability.** The internal cognitive or computational process generating a mathematical result is not directly identical with the linguistic explanation offered after the fact. To preserve this distinction, we use the following terminology.

Definition 3.8. A *reasoning trace* for a trajectory τ is a representation

$$\mathrm{Tr}(\tau) = \langle R_0, R_1, \ldots, R_m \rangle$$

that records selected working units, operations, dependencies, and invariant profiles from τ. The trace is *faithful relative to a goal* when it contains enough information to reconstruct and evaluate every transition materially relevant to the claimed mathematical result.

Faithfulness does not require exhaustive access to every internal activation of a human brain or an artificial neural network. It requires operational adequacy: the trace must make visible the conceptual commitments whose preservation determines whether the solution is valid. Thus, a useful trace is neither an unrestricted stream of tokens nor a polished proof that suppresses all generative structure. It is an intermediate representational object connecting discovery, verification, explanation, and training.

For an LLM, latent states may encode aspects of conceptual organization, but they should not be identified without further evidence with mathematical concepts, mental spaces, or blends. The formal states introduced here are therefore normative and analytical objects. They specify which distinctions a cognitively grounded computational architecture, annotation scheme, or diagnostic probe should attempt to preserve. This caution is especially important when the framework is used to define higher-level loss functions: a metric over latent representations is cognitively meaningful only insofar as its probes reliably track mathematically relevant structures and dependencies.

3.7. **Scope of the framework.** The preceding definitions supply a common interface for the constructions that follow. Section 4 specifies one particularly important class of mathematical working units: formal mathematical concepts with explicit syntactic and semantic components. Formal metaphors can then be understood as controlled correspondences between selected conceptual structures, while conceptual blends are admissible integrations whose output contains emergent structure and remains connected to its inputs through recoverable projections.

Three methodological commitments should be emphasized. First, the framework is representationally plural: symbolic, diagrammatic, linguistic, and model-theoretic forms may all contribute to a conceptual state. Second, it is cognitively constrained but not offered as a complete psychological model.

Its purpose is to formalize operational regularities that are relevant to mathematical creation and computational realization. Third, it is process-sensitive: mathematical competence is evaluated not only by terminal correctness, but also by the coherence, provenance, and strategic organization of the trajectory through which a result is obtained.

These commitments explain why formal conceptual blending is treated below neither as arbitrary combination nor as a merely decorative analogy. It is a structured metamathematical operation acting on conceptually organized inputs, subject to selective projection, semantic viability, invariant-sensitive integration, and recoverable relations to the network from which the blend emerges. The present preliminaries therefore provide the bridge between the general AMI program described in the preceding section and the formal constructions developed in the remainder of the article.

3.8. **Operational Role of the Metamathematical Preliminaries in the Subsequent Models.** The preceding framework is not intended to remain an independent descriptive layer detached from the formal constructions developed below. Its function is to provide a common operational language in which formal mathematical concepts, metaphors, conceptual blends, reasoning trajectories, and cognitively informed training objectives can be represented as different levels of the same mathematical architecture. We now make these connections explicit.

The general relation among the constructions can be summarized schematically as follows. A formal mathematical concept determines a structured mathematical working unit; a metaphor determines a partial transformation between conceptual states; a conceptual blend determines a generative metamathematical operation subject to a specific invariant profile; and a reasoning process involving such operations determines a trajectory whose observable projection can be evaluated by means of the losses and benchmark criteria introduced later. Thus, the notions developed in this section are not additional objects competing with the definitions of Sections 4–7. They specify the state space, transition structure, and observational conditions within which those definitions become computationally operative.

Consider first a formal mathematical concept, as defined in Section 4, of the form

$$C = \langle \mathsf{Log}, \Sigma, A, M \rangle.$$

Such a concept can be embedded into the present metamathematical framework through an associated mathematical description

$$\mathfrak{D}_C = \langle \mathsf{Pres}(\mathfrak{D}_C), \mathsf{Cont}(\mathfrak{D}_C), \mathsf{Role}(\mathfrak{D}_C) \rangle,$$

where

$$\mathsf{Pres}(\mathfrak{D}_C) = (\mathsf{Log}, \Sigma, A), \qquad \mathsf{Cont}(\mathfrak{D}_C) = M,$$

and $\mathsf{Role}(\mathfrak{D}_C)$ records the function performed by C in the current mathematical activity. Depending on the task, the same formal concept may act as an input theory, a generic space, a target structure, a source of examples, an auxiliary construction, or an emerging blend. The fourth component M of the formal concept therefore supplies its semantic content, while the role component introduced in Definition 3.2 makes explicit its changing cognitive function within a reasoning process.

This correspondence explains why the formal definition of a concept alone is insufficient for modeling mathematical activity. The quadruple $\langle \mathsf{Log}, \Sigma, A, M\rangle$ determines what the concept formally is, but it does not determine which parts of it are currently selected, which relations are guiding the reasoning, or why the concept has been introduced. These additional aspects are represented by a conceptual state

$$Q_C = \langle \mathcal{W}_C, \mathcal{F}_C, \mathcal{G}_C, \mathcal{R}_C\rangle.$$

Here $\mathcal{W}_C$ may contain the concept C, selected axioms from A, distinguished models from M, examples, relevant sub-concepts, conjectures, and auxiliary constructions. The focused collection $\mathcal{F}_C$ identifies those components currently active; $\mathcal{G}_C$ specifies the mathematical goal for which they are being used; and $\mathcal{R}_C$ records relations such as logical consequence, inclusion of sub-concepts, satisfaction, structural correspondence, incompatibility, and dependence. In this way, the formal objects of Section 4 acquire an operational position inside a dynamically evolving reasoning process.

The definitions of total, global, and local metaphors in Section 5 instantiate the notion of a partial metamathematical operation. Let

$$\kappa : C_1' \longrightarrow_G C_2'$$

be a local metaphor between suitable sub-concepts of C_1 and C_2. The corresponding operation

$$\Omega_\kappa : \mathcal{Q} \rightharpoonup \mathcal{Q}$$

is defined only on states whose workspace contains the required source and target structures and for which the symbolic correspondence underlying κ is well formed. Its action introduces the translated signature, axioms, and structural relations into the target conceptual environment, modifies the current focus, and may generate new goals or inferential possibilities.

The partiality of Ω_κ is determined concretely by the requirements imposed upon the metaphor. Preservation of the internal taxonomy of symbols contributes to type and arity coherence; the translation of well-formed source sentences contributes to representational well-formedness; and, in the case of a total metaphor, derivability of the translated axioms from the target axioms contributes to semantic viability and dependency preservation. A global metaphor relaxes the last condition, whereas a local metaphor restricts the operation to selected sub-concepts. The distinctions introduced in Section 5 can therefore be recovered as distinctions among the domains, invariant profiles, and preservation requirements of the associated operations.

More precisely, if Q contains the relevant fragments of C_1 and C_2, then an application

$$Q \xrightarrow{\Omega_\kappa} Q'$$

is admissible only when the structural correspondences declared by κ are preserved and when every translated component has an identifiable provenance in the source concept. The conceptual scope of κ becomes part of the workspace of Q', while the role assigned to that scope determines whether it functions as a conjectural analogy, a semantically validated transfer, or an auxiliary structure for subsequent inference. The preliminaries thus provide the machinery needed to distinguish a mathematically controlled metaphor from an accidental symbolic resemblance.

The conceptual blending model of Section 6 constitutes a richer instantiation of the same framework. Let

$$V = \langle G, C_1, C_2, \kappa_1, \kappa_2 \rangle$$

be a V-diagram, and suppose that Q_V is a conceptual state whose workspace contains the three concepts, the two input mappings, and the relations constituting the diagram. A blending operation is then represented as a partial transformation

$$\Omega_{\text{blend}} : Q_V \rightharpoonup Q_B,$$

where the output state contains a candidate blended concept

$$B = \langle \mathsf{Log}, \Sigma_B, A_B, M_B \rangle.$$

The operation is partial because not every pair of conceptual inputs, and not every collection of projections from them, generates a coherent blend.

The six conditions in Definition 6.1 determine the invariant profile

$$\mathcal{I}_{\text{blend},Q_V} = \{\mathcal{I}_{\text{cross}}, \mathcal{I}_{\text{proj}}, \mathcal{I}_{\text{emerg}}, \mathcal{I}_{\text{top}}, \mathcal{I}_{\text{unpack}}, \mathcal{I}_{\text{int}}\}.$$

The partial-cross-space condition specifies the structural correspondences that must remain available between the inputs. Selective projection identifies which components may enter the blend and prevents projection from being confused with indiscriminate union. Emergent structure requires the semantic content of the output to exceed the semantic content already supplied by the input network. The topology condition constrains the compatibility of the projections. Unpacking guarantees the recoverable provenance of the inherited structure, while integration requires the projected material and its semantic closure to constitute a sufficiently unified conceptual object.

Consequently, the definition of an admissible operation given earlier is instantiated here by mathematically explicit conditions. Well-formedness and type coherence constrain the projected signatures and sentences; semantic viability is expressed through the non-emptiness and structural properties of

M_B; dependency preservation is encoded by the topology of the blending network; provenance is enforced by selective projection and unpacking; and goal relevance is determined by the role that the emerging concept plays in the active reasoning problem. The formal definition of blending is therefore a specialized admissibility criterion for a particular family of generative meta-mathematical transformations.

This connection also clarifies the status of emergent structure. Emergence is not identified with arbitrary novelty. A component of B is genuinely emergent only when it is not already contained in the semantic closures associated with the input diagram and nevertheless arises through an admissible integration whose inherited components can be traced back to their sources. Novelty is thereby constrained simultaneously by semantic extension, structural coherence, and recoverable provenance. These are precisely the conditions required for distinguishing conceptual creation from uncontrolled deformation.

The framework becomes dynamic when several metaphors, projections, weakenings, particularizations, or blending operations are composed. A mathematical construction developed through these operations determines a trajectory

$$\tau = \left(Q_0 \xrightarrow{\Omega_0} Q_1 \xrightarrow{\Omega_1} \cdots \xrightarrow{\Omega_{n-1}} Q_n\right).$$

For example, an initial state may contain two concepts and a problem requiring their comparison; a metaphorical operation may identify a common structural fragment; a weakening may suppress irrelevant axioms; a blending operation may construct an integrated concept; and an exemplification may finally produce a model in which the consequences of that concept can be tested. The final proof or mathematical object records only one certified result of this trajectory. The trajectory itself records the conceptual transformations through which that result became available.

This dynamic interpretation supplies the missing formal foundation for the computational model of Section 7. Let

$$h_0, h_1, \ldots, h_T \in H$$

be a latent trajectory generated by an LLM. Since a latent state should not be identified directly with a mathematical concept, we introduce an operation-sensitive probing or decoding map

$$\mathcal{P} : H \longrightarrow \widehat{\mathcal{Q}},$$

where $\widehat{\mathcal{Q}}$ is a space of empirically recoverable approximations to conceptual states. We write

$$\widehat{Q}_t = \mathcal{P}(h_t) = \langle \widehat{\mathcal{W}}_t, \widehat{\mathcal{F}}_t, \widehat{\mathcal{G}}_t, \widehat{\mathcal{R}}_t \rangle.$$

The blended representation $B(h_t)$ used in Section 7 can then be understood as the blend-relevant projection of $\widehat{Q}_t$ rather than as an unqualified identification of the latent vector h_t with a cognitive blend.

Under this interpretation, the local coherence condition introduced later,

$$d\big(B(h_{t+1}), T(B(h_t), \Delta_t)\big) \leqslant \epsilon,$$

is a computational approximation to the requirement that

$$Q_t \xrightarrow{\Omega_t} Q_{t+1}$$

be an admissible and locally coherent metamathematical transition. The transformation T approximates an operation Ω_t from the admissible families defined above, while Δ_t specifies the local reasoning increment or the particular parameters of that operation. The distance d measures the discrepancy between the conceptual state predicted by the declared operation and the state recovered from the model's next latent representation.

The components of the conceptual state also determine the interpretation of the distances and higher-level losses proposed in Section 7. A graph-based distance primarily evaluates changes in $\mathcal{R}$ by testing whether inferential, analogical, or dependency relations have been preserved. A type-consistency distance evaluates the well-formed organization of units in $\mathcal{W}$ and their relations in $\mathcal{R}$. A semantic distance compares the contents and model-theoretic commitments associated with the mathematical descriptions present in successive workspaces. An invariant-preservation distance evaluates the profile $\mathcal{I}_{\Omega,Q}$ governing the transition. A strategy-preservation loss evaluates the stability or legitimate revision of the goal hierarchy $\mathcal{G}$ over longer subtrajectories. Changes in focus $\mathcal{F}$ can additionally be used to detect whether the model has abandoned an essential hypothesis, activated an irrelevant object, or failed to foreground a newly generated proof obligation.

Accordingly, the higher-level objective

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \lambda_1 \mathcal{L}_{\text{blend}} + \lambda_2 \mathcal{L}_{\text{inv}} + \lambda_3 \mathcal{L}_{\text{dep}} + \lambda_4 \mathcal{L}_{\text{strat}}$$

does not introduce an unrelated family of heuristic penalties. Each term evaluates a different component of the metamathematical architecture. The task loss evaluates the terminal result; the blending loss evaluates the local admissibility of conceptual integration; the invariant loss evaluates preservation relative to $\mathcal{I}_{\Omega,Q}$; the dependency loss evaluates the evolution of $\mathcal{R}$; and the strategy loss evaluates coherence in $\mathcal{G}$ across longer reasoning intervals. The hierarchy of losses is therefore derived from the decomposition of conceptual states and transitions established in the present preliminaries.

The same correspondence determines what must be represented in a cognitively grounded benchmark. A benchmark instance should not be restricted to a pair consisting of a problem and its final answer. Its richer form is

$$\mathfrak{B} = \langle x, Q_0, \mathfrak{O}, \mathfrak{I}, \mathrm{Tr}(\tau), y, \mathfrak{E} \rangle,$$

where x is the problem, Q_0 is the initial conceptual state, $\mathfrak{O}$ specifies the admissible families of operations, $\mathfrak{I}$ specifies the invariant profiles that must be preserved, $\mathrm{Tr}(\tau)$ is an observable reasoning trace, y is the expected terminal result, and $\mathfrak{E}$ is an evaluation protocol. The protocol may score terminal correctness, local admissibility, preservation of mathematical objects, management of dependencies, provenance of generated structures, and global strategic coherence.

This formulation does not require unrestricted access to the model's internal reasoning. The benchmark may evaluate an externalized trace, structured annotations, proof states, decoded dependency graphs, or diagnostic probes. What matters is that the selected representation contains sufficient evidence to assess the transitions materially responsible for the mathematical result. The distinction between τ and $\mathrm{Tr}(\tau)$ therefore prevents the framework from assuming that a linguistic chain of thought transparently reproduces the complete internal computation of the model.

We can now state the operational role of the preliminaries in a compact form:

$$\begin{aligned}
\text{formal concepts} &\longmapsto \text{working units and mathematical descriptions},\\
\text{metaphors} &\longmapsto \text{partial transformations of conceptual states},\\
\text{conceptual blends} &\longmapsto \text{admissible generative operations},\\
\text{blending conditions} &\longmapsto \text{operation-specific invariant profiles},\\
\text{reasoning processes} &\longmapsto \text{metamathematical trajectories},\\
\text{model representations} &\longmapsto \text{probed approximations of conceptual states},\\
\text{losses and benchmarks} &\longmapsto \text{evaluations of trajectories and traces}.
\end{aligned}$$

The preliminaries thereby supply the state-space semantics of the later models. Section 4 specifies the principal formal objects populating that space; Section 5 specifies a family of structure-sensitive transformations; Section 6 specifies the admissibility conditions for generative integration; and Section 7 translates states, transformations, dependencies, and invariant profiles into computational observables, training objectives, and benchmark dimensions. In this sense, the remainder of the article is not merely compatible with the metamathematical framework: it constitutes a sequence of increasingly concrete instantiations of it.

4. Formal Mathematical Concepts

Definition 4.1. A *mathematical concept* C is defined as a structured quadruple

$$C = \langle L_{og}, \Sigma, A, \mathbb{M} \rangle,$$

where the components satisfy the following conditions.

- L_{og} denotes a logical system or fixed logic[2].

[2]Here we implicitly assume that L_{og} is sufficiently expressive to provide well-defined notions of syntax, models, satisfiability, and semantic truth.

- Σ represents the complete signature (or alphabet) of symbols used in the formal description of the concept[3].
- A is an explicit collection of well-formed sentences constructed exclusively with symbols from Σ and generated according to the syntactic formation rules prescribed by the logic L_{og}.
- $\mathbb{M}$ denotes the semantic closure of A within the logic L_{og}; that is, $\mathbb{M}$ consists of the class of all models satisfying every sentence contained in A.[4]

Furthermore, given another concept $C' = \langle L_{og}, \Sigma, A', \mathbb{M}' \rangle$, we say that C' is a *sub-concept of* C whenever the set of axioms A' is contained in A. In such a case, C is referred to as a *super-concept* of C'.

5. Formal Metaphorical Reasoning

Let us recall the notion of a mathematical concept introduced above, understood for instance as a many-sorted first-order theory equipped with a specific signature. Within this conceptual framework, we now introduce the notions of *total*, *global*, and *local metaphors* between mathematical concepts.

Definition 5.1. Let

$$C_1 = \{L_{og1}, \Sigma_1, A_1, \mathbb{M}_1\} \quad \text{and} \quad C_2 = \{L_{og2}, \Sigma_2, A_2, \mathbb{M}_2\}$$

be mathematical concepts (for example, many-sorted first-order theories in the category $\mathbb{C}oncepts$[5]).

A *total metaphor* from C_1 to C_2 is a morphism

$$\eta : C_1 \to C_2$$

(also denoted $\eta : C_1 \rightrightarrows_T C_2$) determined by a correspondence

$$\lambda : \Sigma_1 \to \Sigma_2$$

[3]This collection may include non-logical symbols such as variables, constants, and functional symbols; logical symbols; as well as auxiliary syntactic symbols such as parentheses, commas, or operators like $\prod$, $\int$, $\bigcap$, or $\otimes$ that must appear explicitly in the language. These symbols are naturally organized into subclasses according to their functional roles (e.g., logical symbols, auxiliary symbols, variables, constants, functional symbols, and relational symbols).

[4]For cognitive completeness one may assume that the sets A and Σ are sufficiently rich to encode a conceptual substratum of C together with symbols describing its many-sorted configuration. However, this additional requirement is not strictly necessary, since the conceptual substratum can in principle be reconstructed from the previous data when needed. Moreover, in the cognitive meta-generations discussed in Ch.[13], this requirement is not essential.

[5]We do not restrict our deductive or ontological framework to many-sorted first-order logic. The present definition applies to any sufficiently expressive logical environment equipped with its own notions of signatures, sentences, satisfaction, and models. For instance, suitable types of institutions may be employed (see [5]).

that preserves the internal taxonomy of symbols[6].

This correspondence induces a canonical translation of each sentence in A_1 into a collection of well-formed sentences $\eta(A_1)$ expressed within the conceptual environment of C_2. In addition, every sentence in $\eta(A_1)$ must be derivable—either syntactically or semantically—from the axioms of C_2.[7]

A *global metaphor* is a weaker notion. It satisfies all structural requirements of a total metaphor

$$\mu : C_1 \to C_2$$

(denoted $\mu : C_1 \rightrightarrows_G C_2$), except that the axiom-preservation condition is not required; in other words, the translated axioms of C_1 need not necessarily be logical consequences of the axioms of C_2.

Suppose now that

$$C_1' = (\Sigma_1', A_1', \mathbb{M}_1') \quad \text{and} \quad C_2' = (\Sigma_2', A_2', \mathbb{M}_2')$$

are sub-concepts of C_1 and C_2, respectively (that is, $\Sigma_i' \subseteq \Sigma_i$ and $A_i' \subseteq A_i$ for $i = 1, 2$).

A *local metaphor* from C_1 to C_2 (or simply a *metaphor*) is then defined as a global metaphor

$$\kappa : C_1' \rightrightarrows_G C_2'.$$

The *conceptual scope* of κ is the concept

$$(\Sigma_2', A_2' \cup' \kappa(A_1'), \mathbb{M}_3),$$

where $A_2' \cup' \kappa(A_1')$ denotes the union of the sentences of A_2' with the translated sentences $\kappa(A_1')$, and $\mathbb{M}_3$ is the corresponding semantic closure.[8]

Finally, these definitions extend naturally to situations where C_1 and C_2 correspond to exemplifications or generic exemplifications.[9] In particular, if C_2 is an exemplification, the axiom-preserving condition of a total metaphor is required only with respect to the distinguished model described in C_2. In that case, the semantic scope consists solely of this fixed model, and the metaphor is defined only when that model satisfies all translated axioms $\kappa(A_1')$. In all other situations, the axiom-preservation condition is verified with respect to the morpho-syntactic structure of C_2.

---

[6]For example, logical symbols are mapped to logical symbols, functional symbols to functional symbols, and so on.

[7]These requirements generalize the conditions imposed in the definition of morphisms in the category $\mathbb{C}oncepts$ (see Ch. [9]).

[8]In practice we often refer only to the signature and axioms of the conceptual scope, assuming implicitly the associated semantic closure. In such cases we speak of the *syntactic scope* of a metaphor.

[9]All possible combinations are allowed.

The central intuition behind this formalization is that, from a pragmatic perspective, metaphoric reasoning in mathematics typically operates between *localized fragments* of conceptual structures rather than between entire theories. For instance, a mathematician may identify a single operation in an algebraic structure A with another operation in a different structure B, and exploit this identification to derive meaningful inferences regarding a conjecture, without explicitly considering all other defining components of A. This phenomenon reflects a cognitive specialization mechanism of the human mind.

6. Formal Conceptual Blending

In this section we present a refined formalization of conceptual blending that extends the classical V-diagram framework consisting of general mathematical concepts (see [9] for the original formulation).

Let C_1, C_2, and G be mathematical concepts grounded in a common logic L'_{og}. A V-diagram consists of two total metaphors

$$\kappa_1 : C_1 \rightrightarrows_T C_2, \qquad \kappa_2 : C_2 \rightrightarrows_T C_2,$$

together with the three concepts involved, as represented in Fig. 6.1.

(6.1)
$$\begin{tikzcd} C_1 & & C_2 \\ & G \arrow[ul, shift left, "\kappa_1"] \arrow[ul, shift right] \arrow[ur, shift left] \arrow[ur, shift right, "\kappa_2"'] & \end{tikzcd}$$

Definition 6.1. Let

$$V = \langle G, C_1, C_2, \kappa_1 : C_1 \rightrightarrows_T C_2, \kappa_2 : C_2 \rightrightarrows_T C_2 \rangle$$

be a V-diagram defined over a common logic L'_{og}.

A mathematical concept

$$B = \langle L_{og}, \Sigma, A, \mathbb{M} \rangle$$

is called a *conceptual blend for* V (or a blend *between* C_1 *and* C_2 *along* G) if there exist two metaphors

$$\lambda_1 : C_1 \rightrightarrows B, \qquad \lambda_2 : C_2 \rightrightarrows B$$

such that the following conditions hold:

(1) **(Partial cross-space mappings)** There exists a local (total) bi-metaphor $\mu : C_1 \rightrightarrows C_2$ whose conceptual scope generates a concept that is meta-isomorphic to G.[10]
(2) **(Selective projection)** The maps λ_1 and λ_2 are potentially partial metaphors mapping structural components of C_1 and C_2 into the blended concept B.
(3) **(Emergent structure)** The semantic closure $\mathbb{M}$ of B contains at least one model that does not belong to any of the semantic closures associated with the members of the V-diagram.
(4) **(Topology)** The compatibility condition

$$\lambda_1 \circ_{\rightrightarrows} \kappa_1 = \lambda_2 \circ_{\rightrightarrows} \kappa_2$$

holds.
(5) **(Unpacking)** The constitutive concepts of V can be (partially, and adequately for the relevant purposes) reconstructed from B through suitable conceptual weakenings, particularizations (for instance at the level of symbolic units), and generic exemplifications.
(6) **(Integration)** The formal union of the syntactic scopes of λ_1 and λ_2, together with its semantic closure, coincides conceptually with the blended concept B.

(6.2)
$$\begin{tikzcd}
& B & \\
C_1 \arrow[ur, shift left, "\lambda_1"] \arrow[ur, shift right] & \circlearrowleft & C_2 \arrow[ul, shift right, "\lambda_2"'] \arrow[ul, shift left] \\
& G \arrow[ul, shift left] \arrow[ul, shift right, "\kappa_1"'] \arrow[ur, shift left] \arrow[ur, shift right, "\kappa_2"] &
\end{tikzcd}$$

The notion introduced above provides a formal characterization of the minimal structural and governing principles underlying a genuine conceptual blend[11] within our cognitive meta-framework. We deliberately omit an explicit condition corresponding to the *web principle*, since within this framework any conceptual manipulation of the blend that preserves it as a coherent conceptual unit implicitly maintains the structural relations of the network from which the blend originates.

[10]A bi-metaphor is a metaphor equipped with an inverse metaphor with respect to the operation of *metaphorical composition*, denoted by $\circ_{\rightrightarrows}$.
[11]See for example [7].

7. Conceptual Blending and Fine-Tuning of LLMs in Advanced Mathematics

The empirical evidence reported in recent cognitively grounded benchmark studies reveals a structural limitation of current large language models (LLMs): their inability to robustly sustain long-horizon reasoning processes involving the generation, transformation, and stabilization of intermediate mathematical structures [14]. This limitation does not arise primarily from deficiencies in symbolic manipulation, but rather from the absence of mechanisms capable of supporting genuine conceptual synthesis. From the perspective developed throughout this article, such synthesis is intrinsically linked to the operation of conceptual blending as a core generative principle.

Within the broader AMI framework, mathematical reasoning is not reducible to sequences of syntactic transformations, but instead emerges from the controlled interaction of multiple conceptual spaces whose integration yields new structural entities [8], [9]. Consequently, the failures observed in advanced benchmarks—particularly those related to lemma invention, invariant preservation, and adaptive reformulation of conjectures—can be reinterpreted as failures in executing coherent blending processes under increasing conceptual load.

This reinterpretation suggests a shift in the design of fine-tuning strategies. Instead of optimizing LLMs exclusively for input-output accuracy, one should consider training regimes that implicitly or explicitly encode the dynamics of conceptual integration. In this context, fine-tuning must evolve toward a process-sensitive paradigm where the model is guided to produce sequences of intermediate representations that satisfy structural coherence constraints. Such constraints are not purely syntactic but reflect deeper invariants governing the interaction of conceptual components, as systematically analyzed in the AMI program [12]. In particular, one can distinguish three fundamental dimensions that should be incorporated into fine-tuning: a) the preservation of structural invariants across intermediate steps, b) the capacity to reconfigure conceptual dependencies when confronted with novel configurations, and c) the stabilization of emergent blended entities under recursive reasoning.

From a training perspective, this implies that datasets should not only contain final solutions but also cognitively meaningful trajectories of reasoning. These trajectories must encode transitions between conceptual states in a way that reflects their internal dependencies, rather than merely their linguistic realization. In parallel, objective functions should be adapted to penalize incoherent conceptual transitions, even when the final answer is superficially correct. This addresses a key pathology observed in current models, namely the production of correct outputs through internally inconsistent reasoning chains, a phenomenon that becomes particularly evident under stress testing conditions [14].

To make the previous considerations operational, we propose a minimal formal framework capturing blending-aware constraints within the training dynamics of LLMs. Let $\mathcal{X}$ denote the space of input representations, and let $\mathcal{H}$ be the latent space of the model. A reasoning trajectory is modeled as a sequence of latent states $(h_t)_{t=0}^{T}$, where $h_0 = \phi(x)$ for an encoder ϕ and h_T generates the final output.

We interpret each transition $h_t \to h_{t+1}$ as an implicit conceptual transformation. In blending-aware regimes, these transformations should approximate a sequence of structured integrations between latent conceptual components. Let us denote by $\mathcal{B}(h_t)$ the (implicit) blended structure encoded at step t. Then, we require that the trajectory satisfies a coherence constraint of the form:

$$d\big(\mathcal{B}(h_{t+1}), \mathcal{T}(\mathcal{B}(h_t), \Delta_t)\big) \leqslant \epsilon,$$

where $\mathcal{T}$ represents an admissible conceptual transformation, Δ_t encodes the incremental reasoning step, and d is a metric on the space of structured representations.

This condition enforces that each step in the reasoning chain corresponds to a valid transformation of a previously established blended structure, rather than an unconstrained symbolic jump. In practice, $\mathcal{B}$ can be approximated through structured probes over the latent space (e.g., graph-based or type-consistent embeddings), while $\mathcal{T}$ can be learned from annotated reasoning traces.

Based on this principle, we define a blending-consistency loss:

$$\mathcal{L}_{\text{blend}} = \sum_{t=0}^{T-1} d\big(\mathcal{B}(h_{t+1}), \mathcal{T}(\mathcal{B}(h_t), \Delta_t)\big),$$

which can be incorporated into the global training objective as:

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \lambda\, \mathcal{L}_{\text{blend}}.$$

This formulation introduces a structural regularization term that penalizes incoherent conceptual transitions, even when the final output is correct. Importantly, it directly targets the failure modes identified in advanced benchmarks [14], where models exhibit correct answers supported by internally inconsistent reasoning trajectories.

This blending-based measure is, in general, more appropriately applied to local reasoning sub-chains within the global inferential process generated by the LLM. This localized application prevents the introduction of a systematic evaluation bias toward blending-driven reasoning, acknowledging that, from a heuristic and metamathematical standpoint, conceptual blending constitutes only one among several foundational cognitive mechanisms involved in mathematical creative thinking [12].

From a metamathematical perspective, this constraint can be interpreted as enforcing a local preservation of conceptual invariants along the reasoning path, aligning the training dynamics of LLMs with the cognitive mechanisms described in the AMI framework [8], [12]. In particular, it provides a first computational approximation to the principle that mathematical reasoning evolves through controlled transformations of blended conceptual structures, rather than through isolated symbolic operations.

7.1. Hierarchical Blending-Aware Losses for Mathematical Reasoning. The preceding formulation suggests that the fine-tuning of large language models for advanced mathematics should not be understood merely as the optimization of final-answer accuracy. In mathematical practice, the final answer is only the terminal manifestation of a much deeper conceptual trajectory. A proof, a counterexample, a definition, or a conjectural reformulation is generated through a sequence of intermediate conceptual states, each of which must preserve certain structural invariants while allowing controlled transformation, specialization, abstraction, or blending. Consequently, a model that reaches a correct final expression through an incoherent sequence of conceptual transitions should not be regarded as having fully solved the task in the cognitively relevant sense.

This point is especially important in the context of advanced mathematical reasoning. Many benchmark failures do not consist simply in giving the wrong numerical answer. Rather, they involve the progressive degradation of conceptual structure: an invariant is introduced and then silently abandoned; an auxiliary lemma is used outside its domain of validity; a local analogy is extended globally without justification; a geometric object is algebraically translated but its defining constraints are lost; or a proof strategy begins in one conceptual space and ends in another without an admissible bridge. These failures are precisely the kinds of errors that a blending-aware fine-tuning regime is designed to detect and penalize.

The standard task loss L_{task} measures whether the model output matches the desired target. In next-token prediction, this is typically a cross-entropy loss; in supervised mathematical fine-tuning, it may measure the discrepancy between the generated solution and an annotated reference solution. However, L_{task} is usually concentrated near the surface form of the final output. It is therefore insufficient for capturing whether the model has preserved the internal conceptual organization required by the problem.

For this reason, the blending-consistency term introduced above should be interpreted as a first instance of a broader family of higher-level reasoning losses. The essential proposal is that training should impose constraints not only at the final answer level, but also at several higher strata of the reasoning hierarchy. In particular, one may distinguish at least four levels: i) the output level, where the final answer is evaluated; ii) the inferential-step level, where local transitions between consecutive reasoning states are evaluated;

iii) the conceptual-trajectory level, where the stability of mathematical objects, invariants, and dependencies is evaluated across longer subchains; iv) the metamathematical-strategy level, where the global coherence of the proof plan, conceptual blend, or problem-solving architecture is evaluated.

In this hierarchy, the ordinary task loss belongs primarily to the first level. By contrast, the blending-consistency loss belongs to the second and third levels. More ambitious cognitive-computational losses should also act at the fourth level, where one evaluates whether the model has selected and preserved an appropriate mathematical strategy. This is crucial because advanced mathematical intelligence often depends less on isolated symbolic operations than on the disciplined management of conceptual architecture.

Let x be a mathematical problem, and let $(h_t)_{t=0}^{T}$ be the latent reasoning trajectory generated by a model. As above, let $\mathcal{B}(h_t)$ denote the structured blended representation encoded at step t. The local blending-consistency constraint is

$$d(\mathcal{B}(h_{t+1}), T(\mathcal{B}(h_t), \Delta_t)) \leqslant \epsilon.$$

As before, T is an admissible conceptual transformation, Δ_t is the local reasoning increment, d is a distance on structured representations, and ϵ is a tolerance parameter. This inequality says that the next conceptual state must be close to the state obtained by applying an admissible transformation to the previous one. It therefore rules out unjustified symbolic jumps.

A corresponding loss is

$$L_{\text{blend}} = \sum_{t=0}^{T-1} d(\mathcal{B}(h_{t+1}), T(\mathcal{B}(h_t), \Delta_t)).$$

The global training objective may then be written as

$$L = L_{\text{task}} + \lambda L_{\text{blend}},$$

where $\lambda > 0$ controls the strength of the blending-consistency regularization. If $\lambda = 0$, the training regime collapses back to the ordinary task-oriented paradigm. If λ is too large, the model may overfit to a rigid notion of conceptual continuity and lose useful exploratory flexibility. The optimal value should therefore depend on the kind of mathematical task. For formal proof verification, λ should be relatively high; for conjecture generation, analogy search, or exploratory concept invention, λ should be lower, allowing a wider range of admissible conceptual variation.

This formulation is compatible with recent evidence that supervising intermediate reasoning steps can improve mathematical performance. Chain-of-thought prompting showed that intermediate reasoning traces can improve complex reasoning in sufficiently large models [25]. Scratchpad methods similarly demonstrated that forcing models to externalize intermediate computations can improve multi-step algorithmic behavior [22]. Process supervision goes further by assigning feedback to intermediate reasoning steps rather than only to final outcomes [20]. The present framework is aligned with that general direction, but it adds a specifically cognitive-metamathematical constraint: the intermediate steps should not merely be locally acceptable linguistic fragments; they should preserve the structured evolution of mathematical concepts, blends, invariants, and dependencies.

7.2. Possible Formal Definitions of the Distance d. The metric d in the blending-consistency constraint should not be understood as a single universal distance. Different mathematical tasks require different structured representations and therefore different metrics. Nevertheless, several concrete definitions are possible.

First, one may define d as a graph-distance between conceptual dependency graphs. Suppose that each latent state h_t is probed or decoded into a graph

$$G_t = (V_t, E_t),$$

where vertices represent mathematical objects, assumptions, lemmas, invariants, definitions, or intermediate constructions, and edges represent inferential relations such as implication, specialization, dependency, analogy, contradiction, projection, or preservation. Then a simple normalized distance is

$$d(G_t, G_{t+1}) = 1 - \frac{|V_t \cap V_{t+1}| + |E_t \cap E_{t+1}|}{|V_t \cup V_{t+1}| + |E_t \cup E_{t+1}|}.$$

This metric is appropriate when the relevant issue is whether the model has preserved the structural dependencies of the argument. For example, in a proof using the First Isomorphism Theorem, the model must preserve the objects "homomorphism," "kernel," "image," "quotient," and the dependency between normality of the kernel and the construction of the quotient. If the model later treats the quotient as arbitrary, or forgets that the kernel must be normal, the graph-distance increases.

Second, one may define d as a type-consistency distance. In many mathematical failures, the model does not lose a fact but changes the type of an object. For instance, it may treat a group element as a subgroup, a point as a function, a topological space as a metric, or an ideal as a number. Let $\tau_t(o)$ denote the type assigned to object o at step t. Then one may define

$$d_{\text{type}}(h_t, h_{t+1}) = \frac{1}{|O_t|} \sum_{o \in O_t} \mathbf{1}_{\tau_t(o) \neq \tau_{t+1}(o)},$$

where O_t is the set of tracked mathematical objects. This loss penalizes conceptual type drift. It is especially relevant in algebraic geometry, category theory, topology, and logic, where the same symbol may denote objects at different levels unless the conceptual hierarchy is carefully controlled.

Third, one may define d as an invariant-preservation distance. Suppose that the reasoning process is required to preserve a collection of invariants

$$I = \{I_1, \ldots, I_m\}.$$

For example, in a problem involving elliptic curves, such invariants may include nonsingularity, discriminant, torsion constraints, reduction type, or point-counting data over finite fields. In a topological problem, they may include connectedness, orientability, Euler characteristic, or homology groups. Then one may define

$$d_{\mathrm{inv}}(h_t, h_{t+1}) = \sum_{j=1}^{m} w_j \, \mathbf{1}_{I_j(h_t) \neq I_j(h_{t+1})},$$

where w_j measures the importance of invariant I_j. This metric is useful when the task is not merely to transform an expression, but to preserve a mathematical structure under an admissible operation.

Fourth, one may define d semantically, using model-theoretic satisfaction. If $\mathcal{B}(h_t)$ is decoded as a local theory

$$\mathcal{T}_t = (\Sigma_t, A_t, M_t),$$

then the distance between $\mathcal{B}(h_{t+1})$ and $T(\mathcal{B}(h_t), \Delta_t)$ can be measured by the degree to which the axioms, translated axioms, or semantic consequences of one are preserved in the other. A simple version is

$$d_{\mathrm{sem}}(\mathcal{T}, \mathcal{T}') = 1 - \frac{|A_{\mathcal{T}} \cap A_{\mathcal{T}'}|}{|A_{\mathcal{T}} \cup A_{\mathcal{T}'}|},$$

although more sophisticated variants should account for logical consequence rather than syntactic identity. This version is closest to the formal framework of mathematical concepts developed earlier in this article.

7.3. Examples of Admissible Conceptual Transformations. The transformation T represents an admissible evolution of a blended conceptual structure. It is not merely a syntactic rewrite rule. It is a cognitively and mathematically meaningful operation that maps one structured conceptual state into another.

Example 1: Irrationality of $\sqrt{2}$. Consider the classical proof that $\sqrt{2}$ is irrational. At an early stage, the model introduces the assumption

$$\sqrt{2} = \frac{p}{q}, \qquad \gcd(p, q) = 1.$$

The corresponding blended structure $\mathcal{B}(h_t)$ combines the conceptual space of real magnitudes with the conceptual space of rational representation by coprime integers. The reasoning increment is

$$\Delta_t = \text{“square both sides and clear denominators.”}$$

The admissible transformation T sends the current blend to the arithmetic divisibility structure

$$p^2 = 2q^2.$$

A coherent next state $\mathcal{B}(h_{t+1})$ should preserve the fact that p and q are coprime while adding the new divisibility relation. If the model infers that p is even, then writes $p = 2k$, and later concludes that q is also even, the trajectory remains coherent. However, if the model silently drops the coprimality condition or treats p and q as arbitrary real numbers, the distance d should increase.

In this example, an adequate ϵ should be very small, because the proof is rigid. A suitable value might be $\epsilon = 0$ for a symbolic verifier, or $\epsilon = 0.05$ for a neural approximation using graph-based probes. The purpose is not to allow creative deviation, but to enforce the preservation of the contradiction structure.

Example 2: Blending Euclidean and Non-Euclidean Quadrilateral Structures. Consider a Saccheri-type quadrilateral. One input space contains Euclidean quadrilaterals with two right angles and equal lateral sides. Another input space contains a modification in which the summit angles are assumed acute. The blend preserves some structure from the Euclidean input while selectively modifying another part of it. This is not arbitrary deformation: the admissible transformation must preserve the shared quadrilateral structure, the equality of the lateral sides, and the rightness of the base angles, while replacing the summit-angle condition.

Here Δ_t may be

$$\Delta_t = \text{“replace the Euclidean summit-angle condition by the acute-angle hypothesis.”}$$

Then $T(\mathcal{B}(h_t), \Delta_t)$ should produce a blended concept in which the object is still a quadrilateral with the relevant inherited constraints. If the model begins deriving consequences as if the figure were still fully Euclidean, then the blend has collapsed back into the source input. If it treats the figure as an arbitrary quadrilateral with no equal sides or right base angles, then the selective projection has failed. In both cases, d should be large.

This example illustrates why blending-aware supervision differs from ordinary symbolic supervision. The issue is not merely whether a formula is copied correctly. The issue is whether the model preserves the correct pattern of projection, inheritance, and emergence across conceptual spaces.

Example 3: Quotient Construction in Group Theory. Let the task be to prove that if $\varphi : G \to H$ is a group homomorphism, then

$$G/\ker(\varphi) \cong \mathrm{im}(\varphi).$$

A coherent reasoning trajectory must preserve several conceptual dependencies. The kernel is not merely a subset of G; it is a normal subgroup. The quotient $G/\ker(\varphi)$ is only available because of this normality. The map

$$g\ker(\varphi) \mapsto \varphi(g)$$

is well-defined because elements in the same coset differ by an element of the kernel. Thus, the proof is governed by a chain of structural dependencies.

In this case, $\mathcal{B}(h_t)$ may encode the blend between the conceptual space of homomorphisms and the conceptual space of quotient groups. The reasoning increment Δ_t may be

$$\Delta_t = \text{“construct the induced map from cosets to the image.”}$$

The admissible transformation T must generate a state in which well-definedness becomes the central local obligation. If the model jumps directly from the existence of φ to the isomorphism without checking well-definedness, the final answer may be correct but the trajectory is conceptually incomplete. A blending-aware loss would penalize this omission even if L_{task} remains small.

Example 4: Algebraic-Geometric Translation. Suppose a model is asked to translate a geometric condition on an affine variety into an algebraic condition on an ideal. The relevant blend combines a geometric input space of points and subvarieties with an algebraic input space of polynomial rings and ideals. A typical admissible transformation is

$$\Delta_t = \text{“replace geometric containment by ideal-theoretic reverse inclusion.”}$$

Thus, if $V(I) \subseteq V(J)$, the algebraic counterpart is

$$\sqrt{J} \subseteq \sqrt{I}.$$

A coherent trajectory must preserve the contravariance between geometry and algebra. One common model failure is to reverse the inclusion incorrectly, writing $\sqrt{I} \subseteq \sqrt{J}$. This is not merely a symbolic mistake. It is a failure to preserve the blended duality between the geometric and algebraic spaces. A suitable d_{sem} or d_{inv} should strongly penalize this transition.

7.4. Higher-Level Losses Beyond Local Blending Consistency. The local loss L_{blend} captures coherence between adjacent reasoning states. However, advanced mathematical reasoning often requires constraints acting over longer intervals. A model may make locally plausible moves while globally losing the proof strategy. Therefore, one should introduce additional losses higher in the hierarchy of reasoning.

A first possibility is an invariant-trajectory loss:

$$L_{\mathrm{inv}} = \sum_{t=0}^{T} \sum_{j=1}^{m} w_j \, \mathbf{1}_{I_j(h_t) \text{ is violated}}.$$

This term penalizes the violation of invariants that should remain stable throughout a subproof. For example, if a proof assumes throughout that a ring is Noetherian, or that a curve is nonsingular, or that a map is continuous, the model should not silently abandon that condition.

A second possibility is a dependency-coherence loss. Let D_t be the dependency graph of the proof at time t. This graph records which claims depend on which assumptions, lemmas, definitions, or constructions. One may define

$$L_{\mathrm{dep}} = \sum_{t=0}^{T-1} d_{\mathrm{graph}}(D_{t+1}, T_D(D_t, \Delta_t)).$$

This loss is especially relevant for detecting cases where a model invokes a lemma before establishing its hypotheses, or uses a conclusion as if it had already been proved.

A third possibility is a strategy-preservation loss. Let $S(h_t)$ denote the global proof strategy encoded at step t. For example, the strategy may be induction, contradiction, construction of a counterexample, reduction to a known theorem, passage to a quotient, compactness, localization, or categorical universal property. Then one may define

$$L_{\mathrm{strat}} = \sum_{k} d_{\mathrm{strat}}(S_{k+1}, T_{\mathrm{strat}}(S_k, \Delta_k)),$$

where the index k ranges over larger reasoning blocks rather than individual tokens or sentences. This term detects whether the model preserves the intended proof architecture. For instance, if a proof begins by induction but never identifies the induction hypothesis or never proves the induction step, then the strategy has not been preserved.

A more complete objective would therefore have the form

$$L = L_{\text{task}} + \lambda_1 L_{\text{blend}} + \lambda_2 L_{\text{inv}} + \lambda_3 L_{\text{dep}} + \lambda_4 L_{\text{strat}}.$$

This expression should not be interpreted as the final architecture of AMI-oriented fine-tuning. It is rather a minimal formal schema indicating how losses may be distributed across different levels of mathematical cognition. The deeper point is that the training signal must be lifted from the level of the final answer to the level of conceptual process.

This is also consistent with the broader AMI thesis that mathematical intelligence emerges from the controlled interaction of cognitive mechanisms rather than from symbolic manipulation alone [8], [12]. Conceptual blending is one such mechanism, but not the only one. A fully cognitive-computational fine-tuning regime should eventually include additional losses associated with exemplification, generalization, specialization, metaphorical transfer, diagrammatic reasoning, and conceptual substratum, among others.

7.5. Relevance for Benchmarking. The same framework also refines the design of mathematical benchmarks. A benchmark that evaluates only final answers cannot distinguish between genuine mathematical competence and lucky symbolic convergence. In contrast, a cognitively grounded benchmark should annotate and evaluate the intermediate conceptual architecture of a solution. This does not mean that every benchmark must require a fully formal proof. Rather, it means that the benchmark should make explicit which conceptual transitions are essential.

For example, in a task involving a finite-field point count, the benchmark should not only record the final number of points. It should also specify the required conceptual operations: reduction modulo p, identification of the relevant polynomial equation, enumeration or character-sum evaluation, preservation of the field structure, and avoidance of integer-domain reasoning after reduction. In a topology task, the benchmark should not only ask for an Euler characteristic; it should also track whether the model correctly identifies cells, incidences, orientability, and boundary relations. In an algebraic-geometry task, the benchmark should track whether the model preserves the distinction between schemes, varieties, coordinate rings, ideals, radicals, and morphisms.

Thus, blending-aware evaluation is not merely an additional scoring mechanism. It changes the ontology of the benchmark. The object being evaluated is no longer only a pair

$$(\text{problem}, \text{answer}),$$

but a structured trajectory

$$(\text{problem}, \text{conceptual states}, \text{admissible transformations}, \text{answer}).$$

This richer object is better aligned with mathematical practice. Human mathematicians do not usually judge a proof solely by its last line. They judge whether the proof preserves definitions, respects hypotheses, manages dependencies, and introduces new objects in legitimate ways. A benchmark for artificial mathematical intelligence should do the same.

7.6. Final Perspective. The proposed blending-aware framework should be understood as a first computational approximation to a more general principle: mathematical reasoning is a structured evolution of conceptual spaces. Current LLMs often succeed at producing convincing surface-level mathematical language, but they remain fragile when required to preserve deep conceptual organization over long horizons. This fragility becomes visible in advanced benchmarks, where the model must coordinate definitions, examples, invariants, diagrams, algebraic transformations, semantic interpretations, and proof strategies.

By introducing losses at higher levels of the reasoning hierarchy, one can begin to train models not merely to answer, but to reason through stable conceptual transformations. The term L_{blend} is therefore not only a regularizer. It is a prototype for a new kind of cognitive-computational supervision. Its purpose is to penalize incoherent conceptual motion and to reward the disciplined generation, transformation, and stabilization of mathematical structures.

8. General Conclusions and Potential Future Perspectives

8.1. **Seminal General Remarks.** The principal objective of the present work has not been to introduce yet another formal language for mathematical reasoning, nor merely to reinterpret conceptual blending from a mathematical perspective. Rather, our aim has been to establish a first computationally oriented formal framework through which one of the most fundamental mechanisms of human mathematical cognition can begin to be treated as an explicit mathematical object suitable for computational implementation.

Conceptual blending has traditionally been investigated within cognitive science as a domain-general mechanism responsible for the emergence of new conceptual structure through the selective integration of multiple mental spaces [7]. Its explanatory power has been demonstrated across an extraordinarily broad range of human activities, including language, scientific discovery, artistic creativity, social cognition, and mathematical invention. Nevertheless, despite its central role within cognitive science, conceptual blending has remained largely outside the scope of rigorous computational mathematics. Existing formalizations have generally been descriptive, cognitive, or philosophical, but they have not provided a sufficiently explicit mathematical framework capable of supporting computational implementations within artificial mathematical reasoning systems.

The present work attempts to reduce this gap by introducing a coherent mathematical formalization that progressively connects cognitive theory, formal logic, metamathematics, and computational reasoning. Rather than proposing an isolated formal definition, we have developed an integrated hierarchy of mathematical structures that begins with mathematical concepts viewed as formal conceptual entities, continues through mathematically defined metaphorical transformations, culminates in a formal characterization of conceptual blending, and finally illustrates how these structures can be incorporated into computational training and evaluation methodologies for contemporary artificial intelligence systems.

From this perspective, the formal objects introduced throughout this article should not be interpreted merely as abstract mathematical definitions. Instead, they constitute candidate computational primitives. Mathematical concepts become explicit computational entities carrying both syntactic and semantic information. Metaphors become mathematically controlled transformations between conceptual structures. Blends become formally governed mechanisms for generating new mathematical objects through selective conceptual integration. Consequently, the framework developed here provides considerably more than a descriptive language for mathematical creativity: it establishes the first elements of a computational architecture capable of representing, manipulating, and eventually generating mathematical concepts in a cognitively meaningful manner.

This transition from descriptive cognitive theory toward computational formalization represents, in our view, one of the principal conceptual advances of the present work. Historically, cognitive science has been remarkably successful in identifying mechanisms underlying human reasoning, while theoretical computer science has developed increasingly sophisticated computational systems capable of manipulating formal structures. However, the interaction between these two research traditions has remained relatively limited in the specific context of advanced mathematical cognition. The framework proposed here attempts to bridge this separation by providing mathematical definitions that preserve the essential cognitive intuitions of conceptual blending while simultaneously making them accessible to computational implementation.

Another important contribution of the present work lies in the fact that the proposed formalization does not appear in isolation. It naturally fits within the broader Cognitive-Computational Metamathematics (Artificial Mathematical Intelligence) program introduced in [8]. Within this program, conceptual blending is not regarded as an isolated cognitive phenomenon, but as one member of a considerably richer family of fundamental cognitive-computational mechanisms involved in mathematical creation, including exemplification, generalization, specialization, conceptual identification, conceptual decomposition, metaphorical transfer, counterexample construction, conceptual compression, and many others described in the initial AMI taxonomy [12]. Consequently, the present article should be understood not as a complete theory of mathematical cognition, but rather as the first rigorous formal development of one foundational component within a substantially broader cognitive architecture.

The progressive structure of this article reflects precisely this objective. The initial sections revisit the cognitive principles underlying conceptual blending and motivate their relevance for mathematical reasoning. These principles are then embedded into the broader AMI framework, where mathematical intelligence is interpreted as the controlled interaction of multiple cognitive mechanisms rather than as the execution of isolated symbolic transformations. Building upon this conceptual foundation, the article introduces formal mathematical concepts as structured conceptual entities, develops mathematically precise notions of metaphorical reasoning between concepts, and subsequently proposes a formal characterization of conceptual blending itself. Finally, these theoretical developments are connected to current research in artificial intelligence through the proposal of blending-aware fine-tuning and benchmarking methodologies for large language models.

Viewed as a whole, these developments constitute a coherent progression from cognitive theory to computational mathematics. Rather than moving directly from psychological intuition to engineering implementation, the article introduces an intermediate formal layer that makes computational realization mathematically tractable. We believe that such an intermediate layer is indispensable for the future development of cognitively grounded artificial mathematical systems. Without sufficiently explicit formal representations of the cognitive operations involved in mathematical invention, it is difficult to envision computational systems capable of extending mathematics in genuinely creative ways rather than merely reproducing previously observed symbolic patterns.

8.2. **The Metamathematical State-Space as a Foundational Contribution.** A central conclusion of the present work concerns not only the formalization of conceptual blending itself, but also the metamathematical architecture required to make such a formalization operationally meaningful. The objects introduced in Section 3 provide the common state-space semantics underlying the subsequent definitions of formal concepts, metaphors, blends, computational trajectories, loss functions, and benchmark criteria. They should therefore be regarded as one of the principal structural contributions of the article rather than as merely preparatory terminology.

The first consequence of this architecture is a precise separation between object-level mathematical activity and the transformations that reorganize the environment in which such activity occurs. Deriving a consequence from fixed hypotheses, evaluating an expression, or verifying that a structure satisfies a specified property belongs to the object level. By contrast, selecting an example, weakening a hypothesis, changing representation, identifying a relational correspondence, introducing an auxiliary object, or integrating selected material from different conceptual domains changes the conditions under which further deductions become available. These operations are metamathematical in the functional sense developed here.

This distinction leads to an important conclusion regarding mathematical intelligence. A system may be locally competent at formal deduction while remaining globally incapable of reorganizing its mathematical environment in a productive and coherent way. Conversely, many decisive advances in mathematical reasoning arise not from the application of a stronger deductive rule, but from a transformation of the conceptual organization of the problem. Accordingly, mathematical competence cannot be adequately characterized solely by the number, length, or correctness of object-level inferences. It must also include the capacity to construct and control the metamathematical states within which those inferences acquire direction and significance.

The notion of a mathematical working unit captures the entities upon which this extended form of reasoning operates. Symbols, formulas, diagrams, examples, conjectures, models, proof fragments, strategies, and complete concepts may all function as working units, depending on the scale of analysis. This multi-scale character is essential. A mathematical concept may be treated as a single unit during one stage of reasoning and decomposed into axioms, examples, representations, and substructures during another. The framework consequently avoids imposing an artificial universal granularity upon mathematical cognition. Instead, the operative unit is determined relative to the current goal, the representation being used, and the transformation to be performed.

The mathematical description

$$\mathfrak{D} = \langle \mathrm{Pres}(\mathfrak{D}), \mathrm{Cont}(\mathfrak{D}), \mathrm{Role}(\mathfrak{D}) \rangle$$

adds a second indispensable distinction. It separates the explicit presentation of a mathematical entity from its intended content and from the role that it performs in the current reasoning process. This tripartite organization explains why extensionally or semantically equivalent formulations need not be cognitively interchangeable. A factorized polynomial, a geometric diagram, a recurrence written as a matrix action, and a coordinate-free structural description may encode equivalent mathematical content while making radically different relations available to the reasoner.

The distinction between semantic equivalence and cognitive equivalence is therefore not incidental. It indicates that a change of representation should be evaluated not only by whether it preserves truth, but also by whether it preserves or productively reorganizes task-relevant roles, dependencies, and inferential affordances. This observation has direct consequences for both human mathematical practice and artificial reasoning systems. It suggests that representation selection is itself a mathematically evaluable component of problem solving and that a model may fail conceptually even when each isolated reformulation remains formally valid.

These working units and descriptions are organized into conceptual states of the form

$$Q = \langle \mathcal{W}, \mathcal{F}, \mathcal{G}, \mathcal{R} \rangle.$$

The workspace $\mathcal{W}$ specifies which mathematical entities are effectively available; the focus $\mathcal{F}$ identifies which of them are currently active; the goal structure $\mathcal{G}$ determines the purpose and hierarchy governing the activity; and the relational component $\mathcal{R}$ records inferential dependencies, correspondences, incompatibilities, type constraints, identity links, and part–whole relations. This decomposition yields a stronger account of mathematical reasoning than a token sequence, a list of propositions, or an unstructured proof state.

In particular, the state decomposition makes several different kinds of failure distinguishable. A reasoner may lack a necessary object in its workspace, possess it without bringing it into focus, activate an irrelevant object, silently modify the current goal, lose a dependency between a lemma and its hypotheses, confuse analogous structures with identical ones, or combine units whose types are incompatible. These failures can produce the same incorrect terminal answer, but they have different cognitive and computational causes. The proposed state structure therefore supplies a principled basis for diagnostic evaluation rather than merely assigning a uniform penalty to every unsuccessful solution.

The metamathematical operations

$$\Omega : \mathcal{Q} \rightharpoonup \mathcal{Q}$$

provide the transition structure over this state space. Their partiality is one of the most consequential features of the framework. Exemplification, weakening, generalization, metaphorical transfer, identification, decomposition, and conceptual blending cannot be meaningfully applied to arbitrary states. Each operation has representational, logical, semantic, and goal-relative preconditions. A generalization can fail because the selected parameter is not structurally independent; a metaphor can fail because it violates type or arity constraints; a weakening can invalidate a required dependency; and a proposed blend can fail because its projected components admit no coherent semantic integration.

The partial-operation model consequently transforms cognitive mechanisms into objects whose domains of applicability can be studied explicitly. This provides a rigorous alternative to treating mathematical creativity as an unconstrained production of novel associations. A mathematically productive transformation must be generative without being arbitrary, and its legitimacy depends upon the conditions under which it is applied as well as upon the structure of its output.

This principle is made precise through the notions of admissibility and invariant profile. An admissible transition must satisfy the appropriate combination of well-formedness, type and arity coherence, semantic viability, dependency preservation, goal relevance, and recoverable provenance. The invariant profile

$$\mathcal{I}_{\Omega,Q}$$

then records the operation-sensitive constraints that must remain valid, be explicitly discharged, or be legitimately revised during the transition. Rigor is thus represented not as the prohibition of change, but as the controlled preservation and transformation of what matters for the operation being performed.

This formulation supports one of the broader conclusions of the article: mathematical novelty and mathematical rigor should not be treated as opposing properties. Novelty becomes rigorous when it is generated through an admissible transformation whose preserved structure, altered commitments, and emergent consequences can be identified. Conceptual blending is the paradigmatic case developed here, but the same principle applies more generally to abstraction, generalization, conceptual compression, metaphorical transfer, and other mechanisms in the AMI taxonomy. Each family of operations may be characterized through its own preconditions, admissibility requirements, and invariant profiles.

Sequences of these transformations produce metamathematical trajectories

$$\tau = \left(Q_0 \xrightarrow{\Omega_0} Q_1 \xrightarrow{\Omega_1} \cdots \xrightarrow{\Omega_{n-1}} Q_n\right).$$

This trajectory is not identical with the final proof or answer. It may include exploratory examples, provisional correspondences, abandoned strategies, intermediate representations, auxiliary constructions, and failed operations whose diagnostic value disappears from the polished exposition. The final proof is generally a certified projection of a richer process. Consequently, two solutions with the same terminal result may embody substantially different levels of mathematical competence, while an unsuccessful solution may nevertheless exhibit a largely coherent trajectory whose final breakdown is local and identifiable.

This distinction has substantial implications for evaluation. Terminal correctness remains indispensable, but it cannot by itself determine whether a system preserved the mathematical objects, dependencies, goals, and invariants that made the result valid. The appropriate unit of analysis is therefore not only the answer but the structured transition from one conceptual state to another. This conclusion supplies the theoretical foundation for the process-sensitive objectives and benchmarking proposals developed in Section 7.

The reasoning trace

$$\mathrm{Tr}(\tau) = \langle R_0, R_1, \ldots, R_m \rangle$$

provides the observable interface through which such trajectories may be evaluated. Importantly, the trace is neither assumed to reproduce an internal cognitive process exhaustively nor identified with an unrestricted linguistic chain of thought. Its purpose is operational: it should preserve sufficient evidence about working units, transformations, dependencies, goals, and invariant profiles to assess the transitions materially responsible for the mathematical result. This avoids two symmetrical errors: reducing mathematical reasoning to its polished final proof and assuming that every verbalized intermediate step transparently reveals the model's internal computation.

Taken together, the structures of Section 3 determine a unified metamathematical layer:

working units	specify what can be manipulated,
mathematical descriptions	separate presentation, content, and role,
conceptual states	organize workspace, focus, goals, and relations,
partial operations	specify admissible families of transformation,
invariant profiles	govern controlled preservation and revision,
trajectories	represent the evolution of mathematical activity,
reasoning traces	make relevant aspects of that evolution observable.

The later sections instantiate this architecture at increasing levels of specificity. The formal concepts of Section 4 provide a distinguished class of structured working units and mathematical descriptions. The formal metaphors of Section 5 instantiate partial transformations between organized conceptual domains. The blending diagrams and blending conditions of Section 6 specify a generative family of operations together with their characteristic admissibility requirements and invariant profiles. Finally, Section 7 translates conceptual states, transitions, dependencies, invariants, trajectories, and traces into candidate computational representations, hierarchical losses, diagnostic probes, and benchmark dimensions.

The resulting architecture also clarifies what the present article has and has not established. It does not claim that the quadruple Q exhausts human mathematical cognition, that every latent state of an LLM possesses an immediate conceptual interpretation, or that every creative discovery can already be reconstructed through a finite catalogue of operations. The framework is instead normative, extensible, and empirically testable. It identifies distinctions that a cognitively grounded mathematical architecture should attempt to represent and provides a formal basis for investigating whether concrete computational realizations preserve those distinctions.

This qualification is methodologically important. Latent neural representations should not be declared to be concepts, blends, or cognitive states merely because they support successful prediction. Their interpretation requires operation-sensitive probes, structured annotations, controlled interventions, or other empirical evidence demonstrating that the relevant mathematical units, dependencies, goals, and invariants can be recovered. The metamathematical structures introduced here therefore impose a discipline of computational accountability: claims about cognitive organization must be supported by observable preservation and transformation properties.

The principal conclusion is that the article does not merely formalize one mechanism for combining mathematical concepts. It proposes the beginnings of a general state-transition ontology for cognitively structured mathematical activity. Within this ontology, mathematical concepts are not static repositories of axioms and models; they are cognitively addressable structures occupying functional roles within evolving workspaces. Metaphors and blends are not informal similarities or arbitrary unions; they are partial operations with identifiable domains, preservation requirements, and possible failure modes. Mathematical reasoning is not reduced to a terminal proof; it is modeled as an invariant-sensitive trajectory whose relevant structure can be externally represented and evaluated.

This metamathematical layer is what permits the transition from a descriptive theory of mathematical cognition to a computational research program. Without it, the later proposals for blending-aware losses, process-sensitive benchmarking, local UMAAs, and global Artificial Mathematical Intelligence would lack a common account of what evolves during reasoning, which transformations are legitimate, what must remain invariant, and what evidence is required to evaluate the process. With it, these proposals can be understood as different computational realizations of a single foundational principle: advanced mathematical intelligence consists not only in deriving correct consequences, but in governing the coherent evolution of mathematically meaningful conceptual states.

8.3. **Computational Consequences and a New Paradigm for Mathematical Artificial Intelligence.** Beyond its formal contribution, the framework developed throughout this article has significant computational implications. Although the formal definitions introduced here were motivated by questions arising in cognitive science and mathematical creativity, they simultaneously provide computational abstractions that may guide the next generation of artificial mathematical reasoning systems. In particular, they suggest that future mathematical AI systems should not be viewed merely as increasingly powerful symbolic engines or statistical language models, but rather as cognitive-computational architectures capable of manipulating structured mathematical concepts through formally controlled cognitive operations.

One of the principal observations motivating this work is that the limitations exhibited by current large language models on advanced mathematical problems cannot be explained solely by deficiencies in symbolic manipulation or insufficient mathematical knowledge. Recent cognitively grounded benchmark studies reveal that many failures occur despite the model possessing all the local mathematical information necessary to solve the problem correctly [14]. Instead, the breakdown frequently arises from the inability to preserve coherent conceptual organization throughout long reasoning trajectories involving intermediate constructions, auxiliary lemmas, invariant preservation, conceptual reformulation, and the coordinated interaction of multiple mathematical structures.

From this perspective, mathematical reasoning should be regarded not as a sequence of isolated symbolic transformations but as the controlled evolution of structured conceptual states. Consequently, the computational object that deserves supervision during training is not only the final answer but also the internal conceptual trajectory that leads to it. This observation naturally motivates a transition from output-centered optimization toward process-centered optimization.

The blending-aware training framework proposed in Section 7 should therefore be interpreted as an initial computational realization of a considerably broader principle. Rather than optimizing exclusively for final-answer accuracy, future mathematical learning systems should incorporate objective functions acting at multiple levels of the reasoning hierarchy. These levels naturally include the correctness of the final solution, the coherence of intermediate inferential steps, the preservation of conceptual dependencies, the stability of mathematical invariants, the consistency of local and global proof strategies, and, more generally, the controlled evolution of mathematical concepts themselves.

Such a hierarchical view of supervision reflects the actual practice of mathematical research. Human mathematicians rarely evaluate a proof exclusively through its final statement. Instead, they continuously verify whether hypotheses remain active, whether auxiliary constructions preserve their intended properties, whether previously established lemmas are used within their domains of validity, whether conceptual transitions remain justified, and whether the global proof architecture continues to support the original mathematical objective. These higher-order structural constraints are largely absent from current fine-tuning methodologies, which typically assign supervision almost exclusively to surface-level outputs.

The formal notions introduced in this article suggest that these cognitive constraints can be represented mathematically and therefore incorporated into computational optimization procedures. The blending-consistency framework presented in Section 7 constitutes one example of this philosophy, but the same methodology naturally extends to many additional cognitive mechanisms within the AMI taxonomy [12]. Generalization, specialization, exemplification, conceptual decomposition, conceptual identification, counterexample generation, conceptual compression, and other cognitive operations may each give rise to their own families of computational objectives, allowing future training procedures to supervise progressively richer dimensions of mathematical reasoning.

An equally important consequence concerns mathematical benchmarking. Existing mathematical benchmarks generally evaluate the correctness of final answers or, more recently, the plausibility of intermediate reasoning traces. While these approaches represent important advances, they remain largely syntactic. The framework proposed here suggests that future benchmarks should instead evaluate the preservation and evolution of explicit conceptual structures. Rather than asking exclusively whether a model arrives at the correct theorem, proof, or numerical result, cognitively grounded benchmarks should evaluate whether the underlying conceptual transformations satisfy mathematically admissible structural constraints throughout the reasoning process.

Such benchmarks would naturally distinguish between superficial symbolic success and genuine conceptual competence. A model that accidentally produces the correct final answer through internally inconsistent reasoning should not receive the same evaluation as a model whose conceptual trajectory remains structurally coherent from beginning to end. Conversely, models exhibiting mathematically meaningful conceptual evolution, even when failing near the final stages of a difficult proof, provide valuable evidence regarding the maturity of their underlying reasoning mechanisms. Consequently, benchmarking itself becomes a tool for measuring cognitive development rather than merely predictive accuracy.

The computational implications extend beyond evaluation. Because mathematical concepts, metaphors, and blends are explicitly formalized, they become candidates for internal computational representations. Instead of treating latent representations as uninterpretable vectors whose mathematical meaning remains inaccessible, future systems may progressively construct explicit conceptual objects equipped with well-defined syntactic structure, semantic interpretation, admissible transformations, and governing invariants. Such representations would substantially improve interpretability while simultaneously enabling more sophisticated forms of reasoning, verification, debugging, and human-AI mathematical collaboration.

Taken together, these observations suggest a gradual transition toward a new computational paradigm for Artificial Mathematical Intelligence. Rather than viewing intelligence as the capacity to predict mathematically plausible text, the AMI perspective interprets mathematical intelligence as the ability to generate, manipulate, integrate, transform, and stabilize structured mathematical concepts through cognitively meaningful computational operations. Conceptual blending represents one particularly important instance of this broader family of mechanisms, but its formalization also illustrates a more general methodological principle: cognitive operations themselves can become mathematically defined computational objects. Once such operations admit rigorous formal representations, they become amenable not only to theoretical investigation but also to algorithmic implementation, computational optimization, empirical evaluation, and systematic refinement. This observation establishes a direct bridge between cognitive science, formal mathematics, and artificial intelligence, opening a research direction in which future mathematical AI systems may be designed around explicit computational models of mathematical cognition rather than exclusively around statistical language prediction.

8.4. **Toward Local and Global Realizations of the Universal Mathematical Artificial Agent.** The computational implications discussed above naturally suggest a broader research direction extending well beyond the fine-tuning of current large language models. Although the Universal Mathematical Artificial Agent (UMAA) has been presented within the AMI framework as a long-term objective [8], the formal structures introduced throughout the present article provide a concrete pathway toward its progressive computational realization. Rather than attempting to construct a fully general mathematical agent from the outset, the proposed framework supports a modular strategy based on increasingly sophisticated local realizations that may eventually converge toward a unified cognitive-computational architecture.

A natural first step consists in the development of domain-specific or local UMAAs. Each local UMAA would operate within a restricted mathematical discipline while implementing only the subset of cognitive mechanisms required for that domain. For example, a local UMAA devoted to elementary number theory could integrate symbolic computation, finite-field arithmetic, Diophantine reasoning, conjecture generation, example construction, invariant preservation, and blending-aware conceptual transformations specialized to arithmetic structures. Likewise, a local UMAA for abstract algebra could focus on algebraic structures, quotient constructions, universal properties, homomorphisms, categorical reasoning, and structural classification. Similar agents may be envisioned for real analysis, algebraic geometry, topology, combinatorics, mathematical logic, differential geometry, dynamical systems, and numerous other mathematical disciplines.

Such specialization offers several scientific advantages. First, each local UMAA can be evaluated within a mathematically controlled environment where both the underlying conceptual structures and the admissible cognitive mechanisms are comparatively well understood. Second, individual cognitive operations—including conceptual blending, exemplification, specialization, generalization, conceptual decomposition, metaphorical transfer, counterexample generation, and conceptual compression—can be implemented, analyzed, and refined independently before being integrated into more comprehensive architectures. Third, these restricted environments provide ideal experimental laboratories for validating computational implementations of the cognitive taxonomy proposed within the AMI framework [12]. Successes and failures observed during these experiments would directly contribute to refining both the computational models and the underlying cognitive theory.

An equally important consequence concerns the interaction between these local systems and existing mathematical software. The AMI framework does not seek to replace contemporary theorem provers, proof assistants, computer algebra systems, satisfiability solvers, or large language models. On the contrary, each of these technologies possesses highly developed capabilities that naturally complement one another. Interactive theorem provers provide rigorous formal verification; automated theorem provers contribute efficient deductive search; computer algebra systems perform exact symbolic computation; numerical software supplies approximation techniques when appropriate; and modern language models offer flexible heuristic generation and linguistic interaction. The principal contribution of AMI is therefore not the replacement of these systems, but the introduction of a higher cognitive-computational layer capable of coordinating their interaction through mathematically meaningful conceptual operations.

Within such an architecture, conceptual blending occupies a particularly important role because it provides one mechanism by which information originating from heterogeneous mathematical domains may be integrated into coherent conceptual structures. Nevertheless, blending should not be interpreted as the unique organizing principle of the architecture. Rather, it represents one component of a considerably richer cognitive ecosystem whose elements cooperate continuously throughout mathematical discovery. Depending upon the problem under consideration, different cognitive mechanisms may dominate the reasoning process. Some situations require conceptual blending; others depend primarily upon systematic exemplification, carefully controlled specialization, conceptual decomposition, analogical transfer, diagrammatic reasoning, abstraction, or iterative refinement. Consequently, the computational realization of AMI should be understood as the progressive integration of an increasingly complete taxonomy of cognitive-computational mechanisms rather than the implementation of any single operation in isolation.

As these local systems mature, their interaction naturally suggests the emergence of progressively more general mathematical agents. Rather than functioning as isolated domain experts, local UMAAs may exchange conceptual objects, transfer partially developed mathematical structures, communicate through formally defined metaphorical transformations, and collaboratively construct blended conceptual spaces extending across multiple mathematical disciplines. For example, developments in algebraic geometry frequently require sophisticated interactions with commutative algebra, number theory, category theory, topology, and homological algebra. A cognitively grounded architecture capable of coordinating specialized local agents through explicit conceptual transformations would therefore more closely resemble the collaborative behavior of contemporary mathematical research communities than the operation of a monolithic symbolic engine.

This observation suggests that the long-term realization of a global UMAA should be viewed as an emergent cognitive-computational ecosystem rather than as a single self-contained algorithm. Within such an ecosystem, mathematical knowledge would no longer consist merely of collections of formal statements or symbolic derivations, but of dynamically evolving conceptual networks whose components continuously generate, transform, validate, reorganize, and integrate mathematical structures. Mathematical concepts, examples, metaphors, blends, conjectures, proofs, counterexamples, invariants, diagrams, computational experiments, and formal verifications would become interacting computational objects governed by explicit cognitive principles.

The formal developments presented throughout this article may therefore be interpreted as one foundational layer within a considerably broader computational architecture. Mathematical concepts acquire explicit formal representations; metaphorical reasoning becomes mathematically controlled transformation; conceptual blending becomes a formally governed mechanism for generating emergent conceptual structures; and these objects become suitable for computational manipulation. Together, these components begin to establish an intermediate representational layer situated between low-level symbolic computation and high-level mathematical creativity. We believe that the existence of such an intermediate layer is essential for any future artificial system whose objective extends beyond theorem proving toward genuine mathematical co-creation.

Ultimately, the distinction between local and global UMAAs should be viewed not as a distinction between different categories of mathematical agents, but as successive stages in the evolution of a unified research program. Local UMAAs provide experimentally accessible computational realizations of individual cognitive mechanisms within restricted mathematical environments. Their gradual integration, together with continued refinement of the underlying cognitive taxonomy, naturally points toward increasingly comprehensive realizations of Artificial Mathematical Intelligence. In this sense, the present work contributes not only a formal theory of conceptual blending, but also one of the first computational building blocks from which future generations of cognitively grounded mathematical agents may eventually be constructed.

8.5. **Long-Term Research Agenda and Final Remarks.** The developments presented throughout this article should be viewed as one component of a considerably broader scientific program whose ultimate objective extends beyond the formalization of conceptual blending itself. Rather, the long-term vision of Artificial Mathematical Intelligence is to establish a unified cognitive-computational theory capable of explaining, representing, implementing, and eventually extending the principal mechanisms responsible for mathematical creativity. From this perspective, conceptual blending constitutes neither an isolated phenomenon nor a complete explanation of mathematical invention. Instead, it represents one particularly important member of a much larger family of cognitive operations whose systematic computational study remains largely unexplored.

One of the central hypotheses underlying the AMI program is that mathematical intelligence cannot be reduced to symbolic deduction alone. The creation of new mathematical knowledge emerges through the coordinated interaction of multiple cognitive mechanisms operating simultaneously at different conceptual scales. Mathematical concepts are generated, decomposed, specialized, generalized, blended, exemplified, reorganized, abstracted, compressed, and continuously refined throughout the reasoning process. Consequently, a comprehensive theory of artificial mathematical intelligence must ultimately provide computational counterparts for this entire cognitive ecosystem rather than for isolated reasoning procedures.

The formal framework introduced in this article illustrates one possible methodology for achieving this objective. Rather than attempting to model cognition directly through statistical approximations, the approach begins by identifying cognitively meaningful operations, proceeds to formulate mathematically rigorous definitions for these operations, and finally investigates how these formal structures may be incorporated into computational architectures. This methodology naturally combines ideas originating in cognitive science, mathematical logic, category-theoretic thinking, metamathematics, symbolic computation, and artificial intelligence into a unified research direction. We believe that this gradual progression from cognitive observation to formal mathematical representation and finally to computational realization provides a promising methodological paradigm for future investigations in Artificial Mathematical Intelligence.

An important consequence of this perspective is that future advances in mathematical AI should not be measured exclusively by increasing benchmark scores or improving the performance of existing language models. Although such empirical improvements remain important, they represent only one aspect of a considerably deeper scientific objective. Equally significant is the progressive understanding of the cognitive principles that govern mathematical reasoning itself. Every new computational formalization of a cognitive mechanism contributes simultaneously to two complementary goals: it expands our theoretical understanding of mathematical cognition while also enlarging the collection of computational building blocks available for future artificial mathematical systems.

The research program outlined here also suggests a gradual convergence between several disciplines that have historically evolved largely independently. Cognitive science contributes theories describing how mathematical thought is organized. Formal logic provides rigorous languages capable of expressing mathematical structure. Computer science contributes computational architectures capable of executing symbolic and numerical procedures. Artificial intelligence contributes learning mechanisms capable of adapting to increasingly complex mathematical environments. The AMI framework proposes that genuine progress toward artificial mathematical discovery will likely emerge from the systematic integration of these disciplines rather than from further isolated development within any single one of them.

Within this broader context, the New Cognitive Foundations of Mathematics program [8, Part I] acquires particular significance. The formalization of conceptual blending presented here may be interpreted as an illustration of the type of conceptual reconstruction envisioned within that program. Similar investigations may eventually be carried out for many additional mathematical notions, including formal concepts, examples, definitions, proof strategies, abstraction mechanisms, diagrammatic reasoning, conceptual compression, explanatory structures, and numerous other cognitive operations involved in mathematical practice. Collectively, these efforts may contribute to the gradual construction of a computationally grounded cognitive foundation for mathematics itself.

Perhaps the most important long-term implication concerns the possibility of transforming mathematical creativity from an exclusively descriptive subject into a formally investigable computational discipline. Historically, the study of mathematical invention has been dominated either by philosophical reflection or by retrospective analyses of famous discoveries. Although these perspectives remain indispensable, they provide relatively limited guidance regarding computational realization. By contrast, the progressive formalization of elementary cognitive mechanisms opens the possibility of studying mathematical creativity through explicit mathematical objects, formally defined transformations, computational experiments, empirical evaluation, and progressively refined theoretical models. Such a transition would significantly broaden the methodological landscape available for investigating mathematical invention.

Naturally, many important challenges remain unresolved. The present formalization addresses only one cognitive mechanism within a considerably larger taxonomy. The computational realization of conceptual blending itself remains at an early stage. Appropriate internal representations for mathematical concepts, scalable implementations of cognitive operators, efficient interaction between symbolic and neural computation, formal verification of cognitively generated structures, automatic acquisition of conceptual representations, and the integration of multiple cognitive mechanisms into unified reasoning architectures all constitute substantial open research problems. We therefore regard the present work not as a completed theory but as an initial formal step within a long-term research agenda whose full realization will require sustained collaboration across mathematics, cognitive science, computer science, artificial intelligence, philosophy, and related disciplines.

Finally, we believe that the central contribution of this article lies not only in the specific formal definitions that have been introduced, but also in the methodological principle that they exemplify. Cognitive mechanisms underlying mathematical reasoning can themselves become mathematically rigorous computational objects. Once this possibility is accepted, entirely new classes of mathematical software, cognitive-computational architectures, benchmarking methodologies, fine-tuning strategies, and artificial mathematical agents become conceivable. The formalization of conceptual blending presented here therefore represents both a concrete contribution in its own right and an invitation to pursue a broader scientific program aimed at understanding and computationally realizing the mechanisms through which mathematics itself is created.

Ultimately, the long-term aspiration of Artificial Mathematical Intelligence is not merely to construct systems capable of solving increasingly difficult mathematical problems. It is to establish a scientific and computational framework within which artificial systems may participate responsibly, transparently, and rigorously in the continuous expansion of mathematical knowledge itself. We hope that the ideas developed throughout the present work constitute one meaningful step toward that objective.

References

1. Jeremy Avigad and John Harrison, *Formally verified mathematics*, Communications of the ACM **57** (2014), no. 4, 66–75.
2. Clark Barrett, Cesare Tinelli, Haniel Barbosa, Aina Niemetz, Mathias Preiner, Andrew Reynolds, and Yoni Zohar, *Satisfiability modulo theories: A beginner's tutorial*, Proceedings of the 26th International Symposium on Formal Methods (FM 2024), Part II (André Platzer, Kristin Yvonne Rozier, Matteo Pradella, and Matteo Rossi, eds.), Lecture Notes in Computer Science, Springer, 2024, pp. 571–596.
3. Gerd Brewka, *Artificial intelligence—a modern approach by stuart russell and peter norvig*, The Knowledge Engineering Review **11** (1996), no. 1, 78–79.

4. Henri Cohen, *Computational number theory, past, present, and future*, Mathematics Going Forward, Lecture Notes in Mathematics, vol. 2313, Springer, 2022, pp. 561–578.
5. Razvan Diaconescu, *Institution-independent model theory*, Studies in Universal Logic, Birkhäuser Basel, 2008.
6. Răzvan Diaconescu, *Institution-independent model theory*, Studies in Universal Logic, Birkhäuser, Basel, 2008.
7. Gilles Fauconnier and Mark Turner, *Conceptual blending, form and meaning*, Recherches en communication **19** (2003), no. 19, 57–86.
8. D. A. J. Gómez Ramírez, *Artificial mathematical intelligence—cognitive, (meta)mathematical, physical and philosophical foundations*, Springer, Cham, 2020. MR 4225288
9. Danny A. J. Gómez-Ramírez, *Conceptual blending in mathematical creation/invention*, Artificial Mathematical Intelligence, Springer, Cham, 2020, pp. 109–131.
10. ______, *General considerations for the new cognitive foundations' program*, Artificial Mathematical Intelligence, Springer, Cham, 2020, pp. 41–52.
11. ______, *Global introduction to the artificial mathematical intelligence general program*, Artificial Mathematical Intelligence, Springer, Cham, 2020, pp. 1–17.
12. ______, *(initial) global taxonomy of the most fundamental cognitive (metamathematical) mechanisms used in mathematical creation/invention*, Artificial Mathematical Intelligence, Springer, Cham, 2020, pp. 165–198.
13. ______, *Meta-modeling of classic and modern proofs and concepts*, Artificial Mathematical Intelligence, Springer, Cham, 2020, pp. 201–249.
14. Danny Arlen Gómez-Ramírez, *Advanced math benchmarking dataset for large language models*, 2025, Preprint.
15. B. Gonçalves, *Irony with a point: Alan turing and his intelligent machine*, Philosophy of the Media (2023), Interprets Turing's learning machine as a co-creative intelligent system.
16. Danny A. J. Gómez-Ramírez and Yuly A. Herrera-Jaramillo, *Conceptual computation in artificial mathematical intelligence as a paradigm-shifting technique in physics and mathematics*, Journal of Physics: Conference Series, vol. 1672, IOP Publishing, October 2020, p. 012015.
17. Robin Hartshorne, *Algebraic geometry*, vol. 52, Springer Science & Business Media, 2013.
18. George Lakoff and Rafael Núñez, *Where mathematics comes from*, vol. 6, Basic Books, 2000.
19. Zhaoyu Li, Jialiang Sun, Logan Murphy, Qidong Su, Zenan Li, Xian Zhang, Kaiyu Yang, and Xujie Si, *A survey on deep learning for theorem proving*, arXiv preprint arXiv:2404.09939 (2024).
20. Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe, *Let's verify step by step*, International Conference on Learning Representations, 2024.
21. M. Saqib Nawaz, Moin Malik, Yi Li, Meng Sun, and M. Ikram Ullah Lali, *A survey on theorem provers in formal methods*, arXiv preprint arXiv:1912.03028 (2019).
22. Maxwell Nye, Anders Johan Andreassen, Guy Gur-Ari, Henryk Michalewski, Jacob Austin, David Bieber, David Dohan, Aitor Lewkowycz, Maarten Bosma, David Luan, Charles Sutton, and Augustus Odena, *Show your work: Scratchpads for intermediate computation with language models*, arXiv preprint arXiv:2112.00114 (2021).
23. The mathlib Community, *The lean mathematical library*, Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP '20) (New Orleans, LA, USA), 2020, p. 15 pages.
24. Alan M. Turing, *Computing machinery and intelligence*, Mind **LIX** (1950), no. 236, 433–460.
25. Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou, *Chain-of-thought prompting elicits reasoning in large language models*, Advances in Neural Information Processing Systems, vol. 35, 2022, pp. 24824–24837.

VISIÓN REAL COGNITIVA (COGNIVISIÓN) S.A.S. ITAGUÍ, COLOMBIA. & ANANKE INTELLIGENCE INC. ONTARIO, CANADA.
Email address: daj.gomezramirez@gmail.com

CASE WESTERN RESERVE UNIVERSITY, 10900 EUCLID AVE, CLEVELAND, OHIO 44106, USA.
Email address: mark.turner@case.edu