A Mathematical Theory of Understanding

Bahar Taşkesen
University of Chicago, Booth School of Business

March 14, 2026

Abstract

Understanding is not merely the accumulation of information. A signal conveys meaning only to a mind that has the structural capacity to decode it. An explanation that clarifies a concept for one learner may be indistinguishable from noise to another whose prerequisite structure cannot yet support it. This observation suggests that understanding is fundamentally structural: what a learner can grasp depends not only on the signal received, but on the conceptual architecture through which it is interpreted. We describe a mind by a concept space together with axioms and finitary expansion rules that determine when new concepts become accessible once their prerequisites are acquired. These rules, together with the axioms, induce an understanding horizon and a family of reachable concept sets. Under a finite-horizon assumption, this family forms a learning space above the axiom set, that is, an antimatroid. To study the operational consequences of this structure, we model teaching as sequential communication with a latent target concept known to the teacher but unknown to the learner. Instructional signals are then processed through a prerequisite-gated parser: a signal is usable only when its target concept is currently ordered for the learner, and otherwise collapses to a null observation. This mechanism yields two limits on the speed of instruction: a structural limit determined by prerequisite reachability and an epistemic limit determined by target uncertainty. As the learner acquires prerequisites, the instructional channel itself becomes more informative: the statistical experiment through which later instruction is interpreted is refined, in the sense of Blackwell. The framework yields structural thresholds for deterministic targets in fixed-horizon teaching and a linear broadcast penalty across heterogeneous learners. The limits of instruction arise not from how much information is sent, but from whether the learner has the structure required to parse it.

1 Introduction

The value of a piece of information depends on the existence of a mind that can decode it. This is evident in ordinary learning: an explanation means nothing to

a listener who lacks the background to follow it, and a lecture conveys nothing to a student who cannot parse its content. Information that no mind can absorb is, in a precise sense, noise. Understanding therefore cannot be reduced to the accumulation of information. Whether a signal carries usable information is not an intrinsic property of the signal itself, but a relation between the signal and the conceptual structure of the mind that receives it.

Over the past century the cost of producing and distributing information has fallen by orders of magnitude, from printed encyclopedias to digital repositories to, most recently, generative artificial intelligence systems capable of producing explanations, proofs, and worked examples on demand. Yet the cost of absorbing information has not fallen in the same way. Individual fields now reach far beyond what any one person can master, and many of the most consequential problems require cross-domain knowledge whose prerequisite chains are long and intricate. Even when information is generated automatically, it still requires a mind capable of verifying, organizing, and integrating it: a generated treatment protocol still requires a practitioner who understands its logic for reliable deployment. As the supply of information expands, the bottleneck shifts toward the learner: the capacity to absorb, parse, and make use of what is produced.

This paper develops a mathematical framework for a structural component of understanding. We do not attempt to model every feature of cognition, such as analogy, abstraction, forgetting, or semantic interpretation. Instead, we ask a narrower question: given a learner with a fixed prerequisite architecture, which concepts are structurally understandable, through which intermediate states can the learner move, and what limits the speed at which instruction can bring the learner to a target?

Our starting point is a formal model of a *mind*. A mind consists of a concept space together with an axiom set and a family of finitary expansion rules specifying which concepts become accessible once their prerequisites have been acquired. These rules induce an understanding horizon, describing what is reachable in principle from the axioms, and a family of reachable acquired concept sets, describing the intermediate states through which a learner can progress by successive prerequisite-respecting acquisitions. Under a finite-horizon assumption, we show that this reachable family forms a learning space above the axiom set, equivalently an antimatroid, and conversely that every such structure admits a representation by an appropriate mind.

To study the operational consequences of this structure, we model teaching as sequential communication with a latent target concept. The teacher knows the realized target but the learner does not. Instructional signals are filtered through a prerequisite-gated parser: a signal is usable only when its target concept is currently ordered for the learner, and otherwise collapses to a common null observation. The effective learner-side channel is therefore not fixed in advance. It depends on the learner's current acquired state and changes as instruction proceeds. The same raw broadcast may convey usable information to one learner while collapsing to noise for another. We call this phenomenon the *relativity of randomness*.

This state dependence creates two distinct obstacles to fast teaching. The first is *structural*: before a target can be acquired, the learner must move through prerequisite-respecting states until the target becomes currently parseable. The second is *epistemic*: the learner must infer which target the teacher intends. Our main quantitative result combines these two bottlenecks into a general lower bound on teaching time. Expected completion time must clear both a structural barrier, determined by the shortest valid route to the target, and an epistemic barrier, determined by the cumulative usable information that can pass through the learner-side channel. In the baseline model, the structural barrier is often the dominant one, but the information-theoretic layer remains essential for understanding when and how instruction becomes usable.

This perspective leads to several consequences. Acquiring prerequisites does more than add concepts: in the sense of Blackwell, it refines the statistical experiment through which later instruction about the target is interpreted. This structural change has operational implications for teaching. Fixed-horizon teaching problems exhibit discontinuous structural thresholds: for deterministic targets, completion probability jumps from zero to one when the teaching horizon reaches the structural distance to the target concept, implying non-concave returns to instructional time and simple failures of uniform resource allocation. The same structural logic also shapes multi-learner settings. Across heterogeneous learners, a common broadcast curriculum is strictly slower than personalized instruction by a factor linear in the number of learner types.

Related literature. The paper draws on several literatures but differs from each in a specific way.

The combinatorial study of feasible learning states originates with knowledge space theory [Doignon and Falmagne, 1999, 2015], where the family of feasible states is taken as a primitive. Independently, Korte et al. [1991] arrived at the same mathematical structure, antimatroids, from the perspective of combinatorial optimization. We recover this structure from a different starting point: a generative model of a mind specified by axioms and finitary expansion rules. The equivalence (Theorem 2.27) shows that the two viewpoints are formally interchangeable, but the generative formulation connects the combinatorial structure directly to closure, derivability, and the teaching bounds developed in the paper.

Shannon's information theory [Shannon, 1948] studies channels whose input-output relationship is fixed. In our framework, by contrast, the learner's parsing map induces an effective channel whose output alphabet depends on the learner's current acquired state. Blackwell's comparison of experiments [Blackwell, 1951, 1953] provides the natural language for this dependence: we show that the parsed experiment induced by a larger acquired state Blackwell-dominates the one induced by a smaller state.

In computational learning theory, Goldman and Kearns [1995] introduce teaching dimension as a combinatorial measure of how many labeled examples suffice to identify a target concept within a learner class. Our setting is different: the main constraint is not only identification, but whether the learner can parse

target-relevant signals at all given its current prerequisites. More recent work in machine teaching studies settings in which a single teacher must instruct multiple heterogeneous learners with a common teaching sequence; for example, Zhu et al. [2017] show that common teaching can be strictly harder than individualized teaching, and Zhu et al. [2018] survey the broader landscape. Our broadcast impossibility result (Theorem 5.6) differs in the source of the penalty: it is driven by prerequisite-gated decodability and the geometry of reachable acquired states, rather than by differences in algorithmic update rules across learners. The term *curriculum* also appears in machine learning, where it typically refers to the ordering of training examples from easy to hard [Bengio et al., 2009]. There the object being shaped is the optimization trajectory of a parametric model; here it is the sequence of prerequisite-respecting states through which a structured learner can move.

The threshold and allocation results in Section 5 are also related in spirit to models of human-capital accumulation [Becker, 1964, Ben-Porath, 1967, Cunha and Heckman, 2007]. Those models study how current investment affects future skill formation, often through complementarity across stages. Our mechanism is different. In our framework, missing prerequisites create structural thresholds: below the relevant structural depth, completion is impossible regardless of strategy, whereas beyond that threshold completion becomes feasible. The resulting non-smoothness comes from prerequisite-gated decodability rather than from an exogenous production technology. The state-dependent information constraint also connects to rational inattention [Sims, 2003]: both frameworks study limits on usable information, but in rational inattention the bottleneck is imposed through an explicit information cost, whereas here it arises endogenously from the prerequisite structure of the mind.

Notation and conventions. We write $\Delta(\Omega)$ for the set of probability distributions on a finite or countable set Ω. Unless stated otherwise, all logarithms are taken to base 2; accordingly, entropy and mutual information are measured in bits. For a set $\mathcal{S}$, we denote its cardinality by $|\mathcal{S}|$ and its power set by $2^{\mathcal{S}}$. Finally, δ_x denotes the point mass at $x \in \mathcal{S}$.

2 Understanding as a Closure System

What does it mean for a learner to "understand" something? A child who knows addition can follow a multiplication lesson built on repeated addition; one who lacks addition cannot follow that explanation. The same words carry information for one mind and are noise for another. Understanding, in this sense, is not an isolated state but a structured dependency: each concept requires certain prerequisites, and those prerequisites may themselves depend on prior knowledge.

To formalize this idea, we introduce a primitive notion of concept and a nonempty *concept space*, whose elements represent the conceptual units under consideration. A mind is then specified by two objects: a set of axioms and a

set of expansion rules. Axioms are concepts taken as given, requiring no further justification. Each expansion rule states that mastery of a specific finite set of concepts, referred to as its prerequisites, unlocks a new concept. Different minds may share the same concept space yet differ in their axioms or expansion rules. In that case, the order in which concepts become learnable differs, capturing the familiar observation that individuals with different backgrounds require different learning paths.

Given a mind, the expansion rules induce a closure operator: starting from any set of known concepts, iteratively apply every expansion rule whose prerequisites are satisfied until no new concepts are added. The resulting closure operator satisfies extension, monotonicity, and idempotence, the standard closure axioms. These are not merely formal conveniences. Extension encodes that knowledge is never lost by derivation. Monotonicity encodes that knowing more can only enlarge what is derivable. Idempotence encodes that once all consequences have been drawn, further application changes nothing. Any reasonable notion of logical or conceptual consequence must satisfy these properties. The closure framework provides the basic structural language in which the notion of understanding will be formalized in the sections that follow.

We now formalize these ideas using closure operators from order theory.

Definition 2.1 (Concept space)**.** A *concept space* is a nonempty set $\mathcal{C}$ whose elements are *concepts*.

The concept space $\mathcal{C}$ is a modeling primitive: its elements may represent facts ("zebras are animals"), skills ("long division"), propositions ("the fundamental theorem of calculus"), or procedures ("how the simplex method works") at any level of granularity. The framework is invariant to this choice. The modeler selects $\mathcal{C}$ in the same way an economist selects the state space in a decision problem or the type space in a mechanism design model: the choice determines which phenomena the model can express, but the theorems themselves do not depend on the particular interpretation. The concept space $\mathcal{C}$ may be finite or countably infinite. When concepts admit finite descriptions, they can be represented as finite strings over a finite alphabet, and $\mathcal{C}$ can therefore be identified with a subset of that set.

Definition 2.2 (Mind)**.** A *mind* over a concept space $\mathcal{C}$ is a triple $\mathfrak{m} = (\mathcal{C}, \mathcal{A}_\mathfrak{m}, \mathcal{E}_\mathfrak{m})$ where:

(i) $\mathcal{A}_\mathfrak{m} \subseteq \mathcal{C}$ is a set of *axioms*,

(ii) $\mathcal{E}_\mathfrak{m} \subseteq 2^{\mathcal{C}}_{\text{fin}} \times \mathcal{C}$ is a set of *expansion rules*, where $2^{\mathcal{C}}_{\text{fin}}$ denotes the collection of finite subsets of $\mathcal{C}$.

The axioms $\mathcal{A}_\mathfrak{m}$ are the concepts that the mind $\mathfrak{m}$ understands *a priori*: they require no prerequisites. Each expansion rule $(\mathcal{S}, c) \in \mathcal{E}_\mathfrak{m}$ states that if all concepts in the finite set $\mathcal{S}$ are currently understood, then the concept c becomes accessible. The set $\mathcal{S}$ is referred to as the *prerequisites* of c under that rule.

The expansion rules $\mathcal{E}_{\mathfrak{m}}$ describe the cognitive architecture of the mind, that is, the wiring that determines what can be derived from what, rather than propositions explicitly known by the learner. A rule in $\mathcal{E}_{\mathfrak{m}}$ is not assumed to be something the learner can articulate; instead, it specifies which concepts become accessible once the learner has mastered the prerequisites. The *teacher*, by contrast, may or may not know $\mathcal{E}_{\mathfrak{m}}$. A teacher with full knowledge of the learner's rules can tailor instruction to the learner's prerequisite structure, whereas a teacher who is ignorant of the learner's type may have to resort to a common broadcast and can then pay the price of universality (Theorem 5.6).

We impose one structural restriction on $\mathcal{E}_{\mathfrak{m}}$: each prerequisite set is finite. Accordingly, the granularity of $\mathcal{C}$, *i.e.*, what counts as a single concept, should be chosen so that realistic explanations can be modeled using finitely many prerequisites. Beyond this finitarity requirement, the level of granularity is a modeling choice.

Remark 2.3. We do not model logical inconsistency or belief revision. Concepts are treated as abstract units, and understanding refers to accessibility under a prerequisite structure rather than to semantic truth. This is a deliberate modeling choice, analogous to Shannon's separation of the engineering problem of communication from the semantic content of messages. Accordingly, a concept in our framework may represent a true theorem, a useful heuristic, or even a widespread misconception. The theory is invariant to this distinction: the teaching bounds depend only on the dependency structure induced by the prerequisite rules and on the information geometry of the teaching interaction, not on the truth value of the concepts themselves.

Example 2.4 (Two minds learning arithmetic)**.** Let $\mathcal{C} = \{a, b, c, d\}$ with the informal readings $a =$ counting, $b =$ addition, $c =$ spatial arrays, $d =$ multiplication.

Mind 1 (algorithmic learner). Axioms $\mathcal{A}_1 = \{a\}$. The expansion rule set $\mathcal{E}_1$ consists of $\{a\} \Rightarrow b$, $\{b\} \Rightarrow c$, $\{b, c\} \Rightarrow d$. This mind first understands addition from counting, then understands spatial arrays through repeated addition, and finally grasps multiplication once it combines repeated addition with the array representation.

Mind 2 (visual learner). Axioms $\mathcal{A}_2 = \{a\}$. The expansion rule set $\mathcal{E}_2$ consists of $\{a\} \Rightarrow c$, $\{c\} \Rightarrow b$, $\{b, c\} \Rightarrow d$. This mind first understands spatial arrays from counting objects arranged in space, then understands addition by combining arrays, and finally reaches multiplication through the same rule $\{b, c\} \Rightarrow d$.

Both minds in Example 2.4 share the same concept space and the same axioms, and both can eventually derive all four concepts, but the order in which concepts become available differs. A concept that one mind derives early may come late for the other. This is the formal expression of *relativity of understanding*: individuals with different cognitive architectures can arrive at the same body of knowledge through fundamentally different paths. We will revisit this example throughout the paper.

Example 2.4 is about learning mathematics, but the framework applies to any domain in which understanding has prerequisite structure. The next example illustrates this.

Example 2.5 (Two minds learning text editing on a computer)**.** Let $\mathcal{C} = \{t, s, k, e\}$ with the informal readings t = typing text, s = selecting (highlighting) text, k = keyboard shortcuts, e = efficient editing.

Mind 3 (mouse-first). The axiom set is $\mathcal{A}_3 = \{t\}$. The expansion rule set $\mathcal{E}_3$ consists of $\{t\} \Rightarrow s$, $\{s\} \Rightarrow k$, $\{s, k\} \Rightarrow e$. This learner first acquires text selection from typing, then acquires keyboard shortcuts once selection is understood, and finally reaches efficient editing once both selection and shortcuts are available.

Mind 4 (shortcut-first). The axiom set is $\mathcal{A}_4 = \{t\}$. The expansion rule set $\mathcal{E}_4$ consists of $\{t\} \Rightarrow k$, $\{k\} \Rightarrow s$, $\{s, k\} \Rightarrow e$. This learner first acquires keyboard shortcuts from typing, then acquires selection through shortcut-based interaction, and finally reaches efficient editing once both selection and shortcuts are available.

Both minds share the same concept space and the same axiom set, and both can ultimately reach e. However, their prerequisite structures differ: in Mind 3, selection unlocks shortcuts, whereas in Mind 4, shortcuts unlock selection. The final rule $\{s, k\} \Rightarrow e$ is shared, but the paths by which its prerequisites are acquired are different.

The expansion rules admit a combinatorial interpretation. They form a directed hypergraph [Berge, 1984] in which each rule $(\mathcal{S}, c)$ is a hyperedge from the prerequisite set $\mathcal{S}$ to the concept c.

Definition 2.6 (One-step expansion)**.** For a mind $\mathfrak{m}$ and a set $\mathcal{K} \subseteq \mathcal{C}$ of currently known concepts, define the *one-step expansion* by $\Phi_{\mathfrak{m}}(\mathcal{K}) = \mathcal{K} \cup \{c \in \mathcal{C} : \exists \mathcal{S} \subseteq \mathcal{K}$ such that $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}\}$.

For Mind 1 in Example 2.4, start from $\mathcal{K} = \{a\}$. The rule $\{a\} \Rightarrow b$ fires, since $\{a\} \subseteq \{a\}$, and therefore $\Phi_1(\{a\}) = \{a, b\}$. Applying the operator again, the rule $\{b\} \Rightarrow c$ fires, whereas $\{b, c\} \Rightarrow d$ does not, since $c \notin \{a, b\}$. Thus $\Phi_1(\{a, b\}) = \{a, b, c\}$. Applying the operator once more, the rule $\{b, c\} \Rightarrow d$ now fires, so $\Phi_1(\{a, b, c\}) = \{a, b, c, d\}$. A further application produces no new concepts, so $\{a, b, c, d\}$ is a fixed point of Φ_1.

Note that $\Phi_{\mathfrak{m}}$ is *extensive*: by definition, the union in Definition 2.6 includes $\mathcal{K}$ itself, so $\mathcal{K} \subseteq \Phi_{\mathfrak{m}}(\mathcal{K})$ for every $\mathcal{K} \subseteq \mathcal{C}$. We use this property freely throughout. Two further properties of $\Phi_{\mathfrak{m}}$ ensure that repeated application yields a well-defined closure. Monotonicity guarantees that knowledge never shrinks, while preservation of directed unions ensures that no concept appears only "at the limit": whenever a concept is derivable from the union of an increasing family of knowledge states, it is already derivable at some stage in that family.

Lemma 2.7 (Monotonicity)**.** *If* $\mathcal{K} \subseteq \mathcal{K}'$*, then* $\Phi_{\mathfrak{m}}(\mathcal{K}) \subseteq \Phi_{\mathfrak{m}}(\mathcal{K}')$.

Definition 2.8. For a mind $\mathfrak{m}$ and a set $\mathcal{K} \subseteq \mathcal{C}$, the *understanding closure* of $\mathcal{K}$ is the smallest fixed point of $\Phi_{\mathfrak{m}}$ containing $\mathcal{K}$: $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}) = \bigcap\{\mathcal{F} \subseteq \mathcal{C} : \mathcal{K} \subseteq \mathcal{F}$ and $\Phi_{\mathfrak{m}}(\mathcal{F}) = \mathcal{F}\}$. The *understanding horizon* of mind $\mathfrak{m}$ is $\mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$.

For a set map $\Phi_{\mathfrak{m}} : 2^{\mathcal{C}} \to 2^{\mathcal{C}}$, a subset $\mathcal{F} \subseteq \mathcal{C}$ is a *fixed point* if $\Phi_{\mathfrak{m}}(\mathcal{F}) = \mathcal{F}$. Fixed points are partially ordered by set inclusion. Given $\mathcal{K} \subseteq \mathcal{C}$, a fixed point $\mathcal{F}^\star$ is the *least fixed point containing* $\mathcal{K}$ if $\mathcal{K} \subseteq \mathcal{F}^\star$ and $\mathcal{F}^\star \subseteq \mathcal{F}$ for every fixed point $\mathcal{F}$ with $\mathcal{K} \subseteq \mathcal{F}$. In our setting, $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$ is defined as the intersection of all fixed points containing $\mathcal{K}$, hence it is precisely the least knowledge state that contains $\mathcal{K}$ and is closed under all expansion rules of mind $\mathfrak{m}$. In particular, the understanding horizon $\mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ is the least fixed point containing the axiom set $\mathcal{A}_{\mathfrak{m}}$, and therefore represents the set of all concepts *potentially accessible* to $\mathfrak{m}$, that is, the theoretical horizon reachable from its axioms under its expansion rules.

Remark 2.9. It is natural to ask whether a teacher can impart new expansion rules, or how a concept unteachable to a toddler (*e.g.*, abstract algebra) eventually becomes learnable years later. In our framework, methods and techniques that informally feel like "rules", such as the chain rule in calculus or *modus ponens* in logic, are modeled as *concepts* $c \in \mathcal{C}$. They are things a learner can be taught. Once mastered, they serve as prerequisites that unlock downstream concepts via the mind's existing expansion rules. The expansion rules $\mathcal{E}_{\mathfrak{m}}$ and axioms $\mathcal{A}_{\mathfrak{m}}$ themselves are not teachable: they represent the learner's fixed cognitive architecture, sensory baseline, or developmental stage over the timescale of a teaching interaction. A concept is strictly unteachable ($c \notin \mathcal{U}_{\mathfrak{m}}$) when this architecture cannot bridge the gap from the axioms. If a learner eventually grasps a concept that was structurally inaccessible to their earlier self, we model this not as a teaching event, but as *cognitive development*: a transition into a new mind $\mathfrak{m}'$ with richer axioms $\mathcal{A}_{\mathfrak{m}'}$, richer expansion rules $\mathcal{E}_{\mathfrak{m}'}$, or both. Our theory bounds the fundamental limits of *teaching* a fixed architecture; the long-term *development* of the architecture itself is a separate process.

Proposition 2.10 (Existence and characterization)**.** *For any mind* $\mathfrak{m}$ *and any* $\mathcal{K} \subseteq \mathcal{C}$*:*

(i) $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$ *exists and is a fixed point of* $\Phi_{\mathfrak{m}}$*.*
(ii) $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}) = \bigcup_{n=0}^{\infty} \Phi_{\mathfrak{m}}^n(\mathcal{K})$*, where* $\Phi_{\mathfrak{m}}^0(\mathcal{K}) = \mathcal{K}$ *and* $\Phi_{\mathfrak{m}}^{n+1}(\mathcal{K}) = \Phi_{\mathfrak{m}}(\Phi_{\mathfrak{m}}^n(\mathcal{K}))$*.*
(iii) *If* $\mathcal{C}$ *is finite, then* $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}) = \Phi_{\mathfrak{m}}^N(\mathcal{K})$ *for some* $N \leq |\mathcal{C}|$*.*

The existence of a least fixed point in Proposition 2.10 follows from the Knaster-Tarski fixed point theorem [Tarski, 1955]; see also [Aliprantis and Border, 2006] for a textbook treatment. We give a self-contained proof for completeness in Section A.

Proposition 2.11 (Axiomatic characterization of understanding)**.** *For a given mind* $\mathfrak{m}$*, the set* $\mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ *is the unique subset of* $\mathcal{C}$ *satisfying:*

(i) Axioms are understood: $\mathcal{A}_{\mathfrak{m}} \subseteq \mathcal{U}_{\mathfrak{m}}$*.*
(ii) Closure under expansion: *if* $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$ *and* $\mathcal{S} \subseteq \mathcal{U}_{\mathfrak{m}}$*, then* $c \in \mathcal{U}_{\mathfrak{m}}$*.*
(iii) Minimality: $\mathcal{U}_{\mathfrak{m}}$ *is the smallest set satisfying* (i) *and* (ii)*.*

Property (i) of Proposition 2.11 ensures that the axioms belong to $\mathcal{U}_{\mathfrak{m}}$. Property (ii) enforces closure under expansion: whenever all prerequisites of a concept

are already in the set, the concept itself must also belong to the set. Many subsets of $\mathcal{C}$ satisfy (i) and (ii); the entire concept space $\mathcal{C}$ is a trivial example. Property (iii) removes this ambiguity by imposing minimality: $\mathcal{U}_{\mathfrak{m}}$ admits no proper subset that both contains the axioms and is closed under the expansion rules. Together, the three properties determine $\mathcal{U}_{\mathfrak{m}}$ uniquely. In this sense, understanding is completely determined by the axioms $\mathcal{A}_{\mathfrak{m}}$ and the expansion rules $\mathcal{E}_{\mathfrak{m}}$, with no additional degrees of freedom.

2.1 Derivations and Equivalence

The closure $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$ tells us *which* concepts are reachable from $\mathcal{K}$, but not *how* they are reached. A derivation makes the "how" explicit: it is a rooted tree whose nodes represent rule applications and base concepts, showing step by step why a concept lies in $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$. By Lemma A.2, every such tree is finite.

Definition 2.12 (Derivation)**.** A *derivation of concept c from $\mathcal{K} \subseteq \mathcal{C}$ in mind $\mathfrak{m}$* is a well-founded rooted tree whose nodes are labeled by concepts, satisfying:

(i) Every node is either a *base node* or a *rule node*:
- A *base node* is a leaf (no children) labeled by a concept in $\mathcal{K}$.
- A *rule node* is labeled by a concept c' and has children in bijection with a set $\mathcal{S}$ such that $(\mathcal{S}, c') \in \mathcal{E}_{\mathfrak{m}}$, with each child labeled by the corresponding element of $\mathcal{S}$.

(ii) The root is labeled by c.

We write $\mathcal{K} \vdash_{\mathfrak{m}} c$ if such a derivation exists.

Example 2.13 (Derivation trees for the two minds)**.** Continuing Example 2.4, consider the derivation of d (multiplication) from the axiom set $\mathcal{K} = \{a\}$. Figure 1 shows the derivation trees for both minds. In each tree, the *leaves* (bottom nodes, drawn as squares) are concepts from $\mathcal{K}$, which represents the starting knowledge. Each *internal node* (drawn as a circle) is a concept derived by applying one expansion rule to its children (the nodes directly below it). The *root* (top node) is the concept being derived.

Reading each tree bottom-up: Mind 1 derives $a \to b \to c \to d$ (addition before arrays); Mind 2 derives $a \to c \to b \to d$ (arrays before addition). The two derivation trees witness the same conclusion, namely that $d \in \mathrm{cl}_1(\{a\}) \cap \mathrm{cl}_2(\{a\})$, but through different intermediate paths. This provides a concrete instance of mind-relativity.

Derivations provide a constructive counterpart to the closure: if a concept belongs to $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$, there must exist a finite chain of rule applications that produces it. The following theorem confirms that the two characterizations are equivalent, that is, nothing belongs to the closure without a derivation, and every derivation stays within the closure.

Theorem 2.14 (Closure-derivability equivalence)**.** *For any mind $\mathfrak{m}$, any set $\mathcal{K} \subseteq \mathcal{C}$, and any concept $c \in \mathcal{C}$, $c \in \mathrm{cl}_{\mathfrak{m}}(\mathcal{K}) \iff \mathcal{K} \vdash_{\mathfrak{m}} c$.*

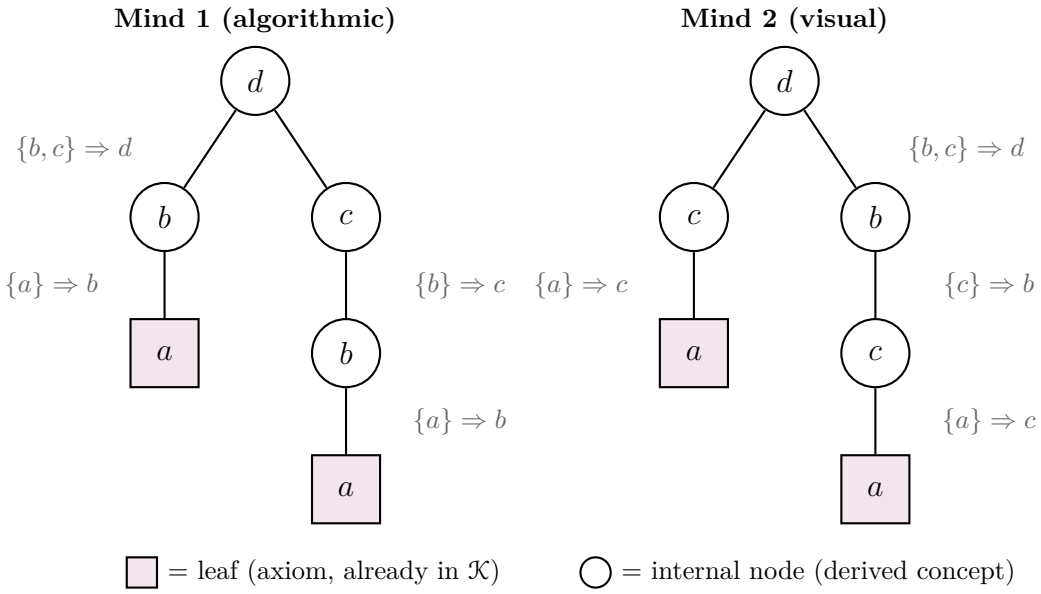


Figure 1: Derivation trees for d (multiplication) from $\mathcal{K} = \{a\}$ (counting) in the two minds of Example 2.4. Each tree is read bottom-up: leaves are concepts already known; each internal node is derived from its children by the expansion rule shown alongside. The root d is the concept being derived. Both trees witness that d belongs to the corresponding understanding closure of $\{a\}$, but through different intermediate paths.

The closure operator $\mathrm{cl}_{\mathfrak{m}}$ induced by a mind satisfies the usual closure axioms (extension, monotonicity, and idempotence). A closure operator that additionally satisfies a finitary property, namely that membership in the closure depends only on finitely many elements, is called algebraic (see Definition A.3).

Theorem 2.15 (Algebraic closure equivalence)**.**

(i) *For any mind* $\mathfrak{m} = (\mathcal{C}, \mathcal{A}_{\mathfrak{m}}, \mathcal{E}_{\mathfrak{m}})$*, the closure operator* $\mathrm{cl}_{\mathfrak{m}} : 2^{\mathcal{C}} \to 2^{\mathcal{C}}$ *is an algebraic closure operator on* $\mathcal{C}$*.*

(ii) *Conversely, for any set* $\mathcal{X}$ *and any algebraic closure operator* $f : 2^{\mathcal{X}} \to 2^{\mathcal{X}}$*, there exists a rule set* $\mathcal{E} \subseteq 2^{\mathcal{X}}_{\mathrm{fin}} \times \mathcal{X}$ *such that, writing*

$$\Psi_{\mathcal{E}}(\mathcal{K}) = \mathcal{K} \cup \{c \in \mathcal{X} : \exists \mathcal{S} \subseteq \mathcal{K} \text{ such that } (\mathcal{S}, c) \in \mathcal{E}\},$$

one has, for every $\mathcal{K} \subseteq \mathcal{X}$*,*

$$f(\mathcal{K}) = \bigcap \Big\{ \mathcal{F} \subseteq \mathcal{X} : \mathcal{K} \subseteq \mathcal{F} \text{ and } \Psi_{\mathcal{E}}(\mathcal{F}) = \mathcal{F} \Big\}.$$

Theorem 2.15 shows that finitary expansion-rule systems and algebraic closure operators are equivalent ways of describing the same finitary consequence relation. In particular, every finitary expansion-rule system induces an algebraic closure operator, and conversely every algebraic closure operator on a set $\mathcal{X}$ admits at least one, generally non-unique, presentation by finitary expansion rules. Thus the rule-based component of a mind should be understood not as additional structure beyond closure, but as a presentation of an algebraic closure operator.

Conceptually, this separates structure from presentation. The expansion rules describe one particular finite-premise decomposition of the underlying consequence relation, while the intrinsic object is the algebraic closure operator itself.

Concretely, let $\mathfrak{X}$ be a nonempty set, let $f : 2^{\mathfrak{X}} \to 2^{\mathfrak{X}}$ be an algebraic closure operator, and let $\mathcal{A} \subseteq \mathfrak{X}$ be a chosen set of axioms. Choose any rule set $\mathcal{E} \subseteq 2^{\mathfrak{X}}_{\text{fin}} \times \mathfrak{X}$ whose induced closure operator is f, as guaranteed by Theorem 2.15(ii). Then $\mathfrak{m} = (\mathfrak{X}, \mathcal{A}, \mathcal{E})$ is a mind whose induced closure operator is f, and whose understanding is $\mathcal{U}_{\mathfrak{m}} = f(\mathcal{A})$.

Thus specifying a mind amounts to specifying an algebraic closure operator together with an axiom set, while the rule formalism provides a finite-premise presentation of that closure structure.

2.2 Ordered and Unordered Information

In classical information theory, the information content of a signal is treated as a property of the source model, independent of the particular receiver. In teaching, however, the usefulness of information is fundamentally relative. The same explanation that substantially reduces uncertainty for a prepared learner may convey little or no usable information to a novice.

This relativity arises because the ability to extract usable information depends on two internal factors: the learner's prerequisite structure $\mathcal{E}_{\mathfrak{m}}$ and the learner's acquired concept set $\mathcal{K}$ at the time of interaction. A concept that is within reach for one mind may be completely inaccessible to another, either because the two minds operate under different prerequisite rules, or because they share the same rules but begin from different acquired concept sets.

Formally, one-step accessibility of a concept is determined by the expansion map: a concept c is reachable from the current acquired concept set $\mathcal{K}$ if and only if $c \in \Phi_{\mathfrak{m}}(\mathcal{K})$. Consequently, the information conveyed by a signal is not determined by the signal alone, but by its position with respect to the learner's mind. This relationship defines the effective channel through which teaching occurs.

Definition 2.16 (Ordered and unordered concept)**.** Let $\mathfrak{m}$ be a mind and let $\mathcal{K} \subseteq \mathcal{C}$ be the set of concepts the learner currently knows. A concept $c \in \mathcal{C}$ is:

(i) *Ordered* for $(\mathfrak{m}, \mathcal{K})$ if $c \in \Phi_{\mathfrak{m}}(\mathcal{K})$. Equivalently, either $c \in \mathcal{K}$, or there exists a rule $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$ such that $\mathcal{S} \subseteq \mathcal{K}$.

(ii) *Unordered* for $(\mathfrak{m}, \mathcal{K})$ if $c \notin \Phi_{\mathfrak{m}}(\mathcal{K})$. Equivalently, $c \notin \mathcal{K}$ and for every rule $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$, at least one prerequisite in $\mathcal{S}$ is missing from $\mathcal{K}$.

At any given stage of the learning process, the set $\mathcal{K}$ represents the concepts *actually acquired so far*. It is *not* assumed to be closed under inference. The closure $\text{cl}_{\mathfrak{m}}(\mathcal{K})$ represents the set of concepts that are *in principle reachable* from $\mathcal{K}$ under the learner's expansion rules. Accordingly, it need not coincide with the learner's current acquired set at a given moment. Later, when we model teaching dynamics (see Section 3), the evolving state $\mathcal{K}_t$ will represent the concepts the

learner has acquired by time t, whereas $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}_t)$ will describe the concepts that are potentially accessible from that state.

Remark 2.17. The distinction between ordered and unordered concepts concerns decodability at the present moment, not whether a signal can be stored for later use. A learner with memory could buffer a signal targeting a currently unordered concept; for example, a student might copy down a formula they do not yet understand. Once the prerequisite concepts enter $\mathcal{K}$, the stored signal may become decodable retroactively. In the memoryless parsing model introduced in Definition 3.3, by contrast, a signal targeting an unordered concept is lost immediately. A natural extension would replace that parser with a delayed-parsing variant, in which raw signals are buffered and re-parsed whenever $\mathcal{K}$ expands. Such a model could lower the teaching-time lower bound, since information presented too early would no longer be wasted.

Definition 2.18 (Valid ordered curriculum)**.** Let $\mathfrak{m}$ be a mind and let $\mathcal{K}_0 \subseteq \mathcal{C}$ be an initial knowledge set. A possibly empty finite sequence $\gamma = ((\mathcal{S}_i, c_i))_{i=1}^{L}$, $L \geq 0$, is a *valid ordered curriculum* starting from $\mathcal{K}_0$ if:

(i) $(\mathcal{S}_i, c_i) \in \mathcal{E}_{\mathfrak{m}}$ for each $i = 1, \ldots, L$;

(ii) defining recursively

$$\mathcal{K}_i = \mathcal{K}_{i-1} \cup \{c_i\}, \quad i = 1, \ldots, L, \tag{1}$$

one has $\mathcal{S}_i \subseteq \mathcal{K}_{i-1}$ for every $i = 1, \ldots, L$.

Definition 2.18 formalizes the idea that a curriculum must respect prerequisites at every step. The rule $(\mathcal{S}_i, c_i)$ can be used only when all concepts in $\mathcal{S}_i$ are already contained in the current set $\mathcal{K}_{i-1}$. Thus the curriculum follows a prerequisite-respecting path, updating the set of concepts acquired by the learner one step at a time. Here $\mathcal{K}_i$ denotes the set of concepts acquired after the first i steps, so that $\mathcal{K}_0 \subseteq \mathcal{K}_1 \subseteq \cdots \subseteq \mathcal{K}_L$.

Theorem 2.19 (Ordering theorem)**.** *For any mind $\mathfrak{m}$ and any target $c^* \in \mathcal{U}_{\mathfrak{m}}$, there exists a valid ordered curriculum $\gamma = ((\mathcal{S}_1, c_1), \ldots, (\mathcal{S}_L, c_L))$, $L \geq 0$, starting from $\mathcal{A}_{\mathfrak{m}}$, such that, if $(\mathcal{K}_i)_{i=1,\ldots,L}$ is constructed as in (1) with $\mathcal{K}_0 = \mathcal{A}_{\mathfrak{m}}$, then $c^* \in \mathcal{K}_L$.*

Example 2.20 (Valid ordered curricula for the two minds)**.** We illustrate Definition 2.18 using the two minds of Example 2.4, both starting from the initial concept set $\mathcal{K}_0 = \{a\}$.

Mind 1 (algorithmic). A valid ordered curriculum for Mind 1 is

$$\gamma_1 = (r_1, r_2, r_3), \qquad r_1 = (\{a\}, b), \quad r_2 = (\{b\}, c), \quad r_3 = (\{b, c\}, d).$$

Writing $c_1 = b$, $c_2 = c$, $c_3 = d$, and defining

$$\mathcal{K}_0^{(1)} = \{a\}, \qquad \mathcal{K}_i^{(1)} = \mathcal{K}_{i-1}^{(1)} \cup \{c_i\} \quad (i = 1, 2, 3),$$

we obtain

$$\mathcal{K}_1^{(1)} = \{a, b\}, \qquad \mathcal{K}_2^{(1)} = \{a, b, c\}, \qquad \mathcal{K}_3^{(1)} = \{a, b, c, d\}.$$

Indeed, at each step the prerequisite set of the selected rule is contained in the current acquired concept set.

Mind 2 (visual). A valid ordered curriculum for Mind 2 is

$$\gamma_2 = (r'_1, r'_2, r'_3), \qquad r'_1 = (\{a\}, c), \quad r'_2 = (\{c\}, b), \quad r'_3 = (\{b, c\}, d).$$

Writing $c'_1 = c$, $c'_2 = b$, $c'_3 = d$, and defining

$$\mathcal{K}_0^{(2)} = \{a\}, \qquad \mathcal{K}_i^{(2)} = \mathcal{K}_{i-1}^{(2)} \cup \{c'_i\} \quad (i = 1, 2, 3),$$

we obtain

$$\mathcal{K}_1^{(2)} = \{a, c\}, \qquad \mathcal{K}_2^{(2)} = \{a, b, c\}, \qquad \mathcal{K}_3^{(2)} = \{a, b, c, d\}.$$

Again, each rule is applicable when used.

Thus the two minds admit different valid ordered curricula from the same starting set. In particular, their first steps must differ. For Mind 1 the only rule whose prerequisite set is contained in $\{a\}$ is $(\{a\}, b)$, whereas for Mind 2 the only such rule is $(\{a\}, c)$. This suggests that a single common curriculum cannot in general respect the structural requirements of both minds simultaneously, foreshadowing the impossibility result of Section 5.2.

Proposition 2.21 (Curricula stay inside understanding horizon)**.** *Let $\mathfrak{m}$ be a mind, and let $\gamma = ((\mathcal{S}_i, c_i))_{i=1}^{L}$ be a valid ordered curriculum starting from $\mathcal{A}_\mathfrak{m}$. Let $(\mathcal{K}_i)_{i=1,\dots,L}$ be constructed as in (1) with $\mathcal{K}_0 = \mathcal{A}_\mathfrak{m}$. Then $\mathcal{K}_i \subseteq \mathcal{U}_\mathfrak{m}$ for every $i = 0, 1, \dots, L$. In particular, if $c^* \notin \mathcal{U}_\mathfrak{m}$, then no valid ordered curriculum starting from $\mathcal{A}_\mathfrak{m}$ can reach c^*.*

Proposition 2.21 draws a boundary around what any curriculum can achieve. If a concept does not belong to the understanding horizon $\mathcal{U}_\mathfrak{m}$, then no sequence of rule applications, however long or carefully arranged, can produce it. The barrier is structural, not epistemic: it is not that the teacher lacks information or that the curriculum is poorly designed, but that the expansion rules of the learner do not connect the axioms to the target concept. In this sense, the understanding horizon $\mathcal{U}_\mathfrak{m}$ is the theoretical horizon of the mind $\mathfrak{m}$.

A concrete illustration is the attempt to convey the visual experience of the color purple to a learner who has been blind from birth. Here the target concept is not the word purple or its descriptive use, but the sensory concept associated with its visual appearance. Such a learner may understand many relational facts about color: that purple is classified between blue and red in certain systems, that particular objects are called purple by sighted speakers, or that light associated with purple occupies a certain range of wavelengths. But if the learner's mind contains no rule path from its existing concepts to that sensory target, then no curriculum, however long or ingeniously ordered, can reach it.

2.3 Reachable acquired concept sets

The closure operator $\mathrm{cl}_{\mathfrak{m}}$ identifies what is reachable in principle from the axiom set, but it does not describe the intermediate concept sets through which a learner may pass on the way to that horizon. This distinction is structurally important and closely related to a central idea in the literature on knowledge spaces, where one studies not only which concepts are ultimately attainable, but also which intermediate learning states are feasible along a learning process [Doignon and Falmagne, 1999, Korte and Lovász, 1983]. Closure is a global notion: if a concept lies in $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$, then it is eventually reachable from $\mathcal{K}$, but it need not already belong to the current acquired concept set $\mathcal{K}$. In particular, closure alone does not record which subsets of $\mathcal{U}_{\mathfrak{m}}$ can arise by successive prerequisite-respecting acquisitions, one concept at a time.

For the structural theory of teaching, we therefore need a finer object than the understanding horizon alone. We introduce the family of *reachable acquired concept sets*: those subsets of $\mathcal{U}_{\mathfrak{m}}$ that can be built from the axioms by a finite sequence of locally valid acquisitions. This family is the natural state space for teaching dynamics. Later we show that, under a finite-horizon assumption, it has the combinatorial structure familiar from the knowledge-space literature: after shifting by the axiom core, it forms an *antimatroid*, equivalently, a learning space. Thus the framework does not take feasible learning states of [Doignon and Falmagne, 1999] as primitive; rather, it derives them from axioms and expansion rules of a mind.

Definition 2.22 (Reachable acquired concept sets)**.** A set $\mathcal{K} \subseteq \mathcal{U}_{\mathfrak{m}}$ is *reachable* if there exists a finite chain $\mathcal{A}_{\mathfrak{m}} = \mathcal{K}_0 \subset \mathcal{K}_1 \subset \cdots \subset \mathcal{K}_L = \mathcal{K}$ such that for each $i = 0, \ldots, L-1$, $\mathcal{K}_{i+1} = \mathcal{K}_i \cup \{c_i\}$, $c_i \in \Phi_{\mathfrak{m}}(\mathcal{K}_i) \setminus \mathcal{K}_i$. Any such chain is called a *witnessing chain* for the reachability of $\mathcal{K}$. Define the *reachable family* of mind $\mathfrak{m}$ $\mathbb{K}_{\mathfrak{m}} = \{\mathcal{K} \subseteq \mathcal{U}_{\mathfrak{m}} : \mathcal{K} \text{ is reachable from } \mathcal{A}_{\mathfrak{m}}\}$. We write $\underline{\mathcal{K}} = \mathcal{A}_{\mathfrak{m}}$, $\overline{\mathcal{K}} = \mathcal{U}_{\mathfrak{m}}$, for the minimum and maximum reachable acquired concept sets.

Assumption 2.23 (Finite understanding horizon)**.** *The understanding horizon* $\mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ *is finite.*

Assumption 2.23 is imposed only to place the reachable family within the finite combinatorial framework of learning spaces and antimatroids. The definition of reachability itself does not require finiteness.

The reachable family $\mathbb{K}_{\mathfrak{m}}$ will later serve as the state space for the teaching dynamics in Section 3, so its internal structure is central to the theory. The next proposition shows that this family has three basic features. It has a distinguished minimum state, every non-minimal reachable state can be obtained from another reachable state by adding a single concept, and it is closed under unions. These properties are natural from the perspective of learning: one can build feasible states step by step, and compatible partial acquisitions can be combined. They also place the reachable family in close correspondence with the combinatorial objects studied in the literature on learning spaces [Doignon and Falmagne, 1999] and antimatroids [Korte and Lovász, 1983].

Proposition 2.24 (Structure of the reachable family)**.** *Under Assumption 2.23, the family $\mathbb{K}_{\mathfrak{m}}$ is finite and satisfies:*

(i) $\underline{\mathcal{K}} = \mathcal{A}_{\mathfrak{m}}$ *is the minimum element of the partially ordered set* $(\mathbb{K}_{\mathfrak{m}}, \subseteq)$*;*
(ii) *for every* $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}}$ *with* $\mathcal{K} \neq \mathcal{A}_{\mathfrak{m}}$*, there exists* $\mathcal{K}' \in \mathbb{K}_{\mathfrak{m}}$ *such that* $\mathcal{K}' \subset \mathcal{K}$ *and* $|\mathcal{K} \setminus \mathcal{K}'| = 1$*;*
(iii) *if* $\mathcal{K}, \mathcal{K}' \in \mathbb{K}_{\mathfrak{m}}$*, then* $\mathcal{K} \cup \mathcal{K}' \in \mathbb{K}_{\mathfrak{m}}$*;*
(iv) $\overline{\mathcal{K}} = \mathcal{U}_{\mathfrak{m}}$ *is the maximum element of* $(\mathbb{K}_{\mathfrak{m}}, \subseteq)$*;*
(v) *ordered by inclusion,* $(\mathbb{K}_{\mathfrak{m}}, \subseteq)$ *is a finite join-semilattice, and for every* $\mathcal{K}, \mathcal{K}' \in \mathbb{K}_{\mathfrak{m}}$ *the join is given by* $\mathcal{K} \vee \mathcal{K}' = \mathcal{K} \cup \mathcal{K}'$*.*

Properties (i) through (iii) identify the core combinatorial features of the reachable family: a distinguished minimum state, one-step accessibility, and union-closure. These are precisely the ingredients that connect the reachable family to the notions of learning space [Doignon and Falmagne, 1999] and antimatroid [Korte and Lovász, 1983].

To make the connection precise, we recall both concepts and their equivalence. An antimatroid on a finite set $\mathcal{E}$ is a family $\mathcal{F} \subseteq 2^{\mathcal{E}}$ satisfying: (i) $\varnothing \in \mathcal{F}$; (ii) for every nonempty $\mathcal{S} \in \mathcal{F}$, there exists $x \in \mathcal{S}$ such that $\mathcal{S} \setminus \{x\} \in \mathcal{F}$ (accessibility); and (iii) $\mathcal{F}$ is union-closed. While matroids [Whitney, 1935] axiomatize independence structures where feasibility is closed downward: every subset of a feasible set is feasible; antimatroids [Korte and Lovász, 1983] capture the complementary pattern: feasibility is closed upward under unions, modeling sequential construction under precedence constraints. Independently, [Doignon and Falmagne, 1999] arrived at the same mathematical structure from a different motivation: modeling the feasible knowledge states of a human learner. They called the resulting object a learning space [Doignon and Falmagne, 2015, Theorem 7], [Doignon and Falmagne, 2016], which is an antimatroid. The standard definition of a learning space takes the empty set as the minimum element, modeling a learner who begins with no knowledge. In our setting the learner starts from the axiom set $\mathcal{A}_{\mathfrak{m}}$, so we introduce a shifted variant that replaces $\varnothing$ with $\mathcal{A}$.

Definition 2.25 ($\mathcal{A}$-based learning space)**.** Let $\mathcal{U}$ be a finite set and let $\mathcal{A} \subseteq \mathcal{U}$. A family $\mathbb{F} \subseteq 2^{\mathcal{U}}$ is called an *$\mathcal{A}$-based learning space* if:

(i) $\mathcal{A} \in \mathbb{F}$ and every $\mathcal{K} \in \mathbb{F}$ satisfies $\mathcal{A} \subseteq \mathcal{K}$;
(ii) for every $\mathcal{K} \in \mathbb{F}$ with $\mathcal{K} \neq \mathcal{A}$, there exists $x \in \mathcal{K} \setminus \mathcal{A}$ such that $\mathcal{K} \setminus \{x\} \in \mathbb{F}$;
(iii) $\mathbb{F}$ is union-closed.

Equivalently, the shifted family $\widehat{\mathbb{F}} = \{\mathcal{K} \setminus \mathcal{A} : \mathcal{K} \in \mathbb{F}\} \subseteq 2^{\mathcal{U} \setminus \mathcal{A}}$ is an antimatroid.

Corollary 2.26 (Shifted antimatroid structure)**.** *Under Assumption 2.23, the reachable family $\mathbb{K}_{\mathfrak{m}}$ is an $\mathcal{A}_{\mathfrak{m}}$-based learning space. Equivalently, the shifted family $\widehat{\mathbb{K}}_{\mathfrak{m}} = \{\mathcal{K} \setminus \mathcal{A}_{\mathfrak{m}} : \mathcal{K} \in \mathbb{K}_{\mathfrak{m}}\} \subseteq 2^{\mathcal{U}_{\mathfrak{m}} \setminus \mathcal{A}_{\mathfrak{m}}}$ is an antimatroid.*

The next theorem characterizes the reachable families generated by minds: they are precisely the $\mathcal{A}$-based learning spaces.

Theorem 2.27 (Representation of reachable families)**.** *Let $\mathcal{C}$ be a finite set, let $\mathcal{A} \subseteq \mathcal{C}$, and let $\mathbb{F} \subseteq 2^{\mathcal{C}}$. The following are equivalent:*

(i) $\mathbb{F}$ *is an* $\mathcal{A}$*-based learning space;*
(ii) *there exists a mind* $\mathfrak{m} = (\mathcal{C}, \mathcal{A}, \mathcal{E}_\mathfrak{m})$ *whose reachable family satisfies* $\mathbb{K}_\mathfrak{m} = \mathbb{F}$. *Moreover, when* (i) *holds, the mind* $\mathfrak{m}_\mathbb{F} = (\mathcal{C}, \mathcal{A}, \mathcal{E}_\mathbb{F})$ *with rule set* $\mathcal{E}_\mathbb{F} = \{(\mathcal{S}, c) : \mathcal{S} \in \mathbb{F},\ c \in \mathcal{C} \setminus \mathcal{S},\ \mathcal{S} \cup \{c\} \in \mathbb{F}\}$ *satisfies* $\mathbb{K}_{\mathfrak{m}_\mathbb{F}} = \mathbb{F}$.

Figure 2 illustrates the reachable family for a mind with axiom set $\mathcal{A}_\mathfrak{m} = \{a\}$ and expansion rules $\{a\} \Rightarrow b$, $\{a\} \Rightarrow c$, $\{b, c\} \Rightarrow d$. Starting from $\underline{\mathcal{K}} = \{a\}$, the learner can acquire b or c in either order, since both are individually unlocked by the axiom. However, d becomes reachable only once both b and c have been acquired, so $\{a, b, c\}$ is the unique gateway to d. The set $\{a, b, d\}$, for instance, does not belong to $\mathbb{K}_\mathfrak{m}$ because the rule for d requires c, which is absent. The figure makes both accessibility and union-closure visible.

Theorem 2.27 characterizes the reachable families generated by minds as the $\mathcal{A}$-based learning spaces. This has two consequences for the present work. First, the feasible knowledge states of a mind need not be postulated as a primitive; they are derived from axioms and expansion rules, and the resulting state space automatically inherits the rich combinatorial structure of an antimatroid. Second, the converse direction guarantees that the mind formalism is fully expressive: any learning space one might wish to study can be generated by a suitably chosen mind. Thus the structural and the generative viewpoints are equivalent. We note, however, that not every union-closed family above the axioms qualifies as an $\mathcal{A}$-based learning space. Accessibility is an additional requirement. It rules out degenerate state spaces in which the learner cannot progress one concept at a time; see Corollary B.1.

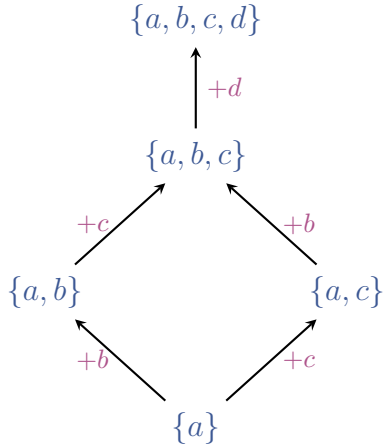


Figure 2: The reachable family $\mathbb{K}_\mathfrak{m}$ for a mind with axiom set $\mathcal{A}_\mathfrak{m} = \{a\}$ and expansion rules $\{a\} \Rightarrow b$, $\{a\} \Rightarrow c$, $\{b, c\} \Rightarrow d$. The concept d becomes reachable only at $\{a, b, c\}$, where both prerequisites are present. Sets such as $\{a, b, d\}$ are structurally unreachable.

3 Teaching and Learning Dynamics

Understanding characterizes which concepts are in principle accessible under a prerequisite structure. Teaching introduces a second challenge beyond accessibility: the learner must identify the teaching target. A signal about addition in a mathematics course, for example, may indicate that addition is itself the intended endpoint, or it may be an intermediate step on the way to multiplication. This is the identification component of teaching.

It is here that intentionality enters. A teaching move is not merely the presentation of a concept; it is an action chosen in light of a target and interpreted by the learner as evidence about that target. To represent this *as*ymmetry, we

model the target concept as a latent variable known to the teacher and unknown to the learner, and we represent the learner's evolving belief as a probability distribution over candidate targets.

The latent target need not be interpreted only as the teacher's intended endpoint. It may also be read as the higher-level concept that renders the currently acquired material globally coherent. On this interpretation, learning involves two coupled dimensions: the acquired concept set expands, while the learner simultaneously infers which larger target those concepts are organizing toward. A concept may therefore be acquired locally before its place in the larger conceptual graph is understood. For example, a learner may acquire many concepts from electromagnetism and electronics while still lacking the bridge concept that connects them to wireless communication. Once that target is identified, previously disconnected material becomes integrated as part of a single explanatory structure.

Teaching dynamics therefore involve both structural and epistemic progress. Structural progress is governed by the prerequisite structure: once the learner is at an acquired set from which a concept is *ordered*, and the appropriate signal is successfully parsed, that concept enters the learner's acquired concept set. Epistemic progress, by contrast, concerns the gradual resolution of uncertainty about the target. Because the learner does not know which target the teacher intends, each signal must play a dual role: it must be a valid instructional step in the prerequisite structure, and it must simultaneously provide evidence that distinguishes the intended target from the alternatives. From the learner's perspective, the observed signal is therefore a random variable whose distribution depends on both the unknown target and the teaching strategy. Each round can convey only a bounded amount of usable information about the latent target, and the total teaching time is governed by the rate at which this epistemic uncertainty is resolved. If the learner knew the target from the outset, the epistemic dimension would disappear and teaching would reduce to the purely structural problem of reaching a known target by a valid curriculum.

We now make these ideas concrete by introducing a stochastic model of teaching.

3.1 A Stochastic Model of Teaching

Fix a probability space $(\Omega_0, \mathcal{F}, \mathbb{P})$ on which all random variables below are defined. Let $\Omega \subseteq \mathcal{C}$ be a finite set of *target concepts*. Let $\Theta : \Omega_0 \to \Omega$ be an Ω-valued random variable representing the realized target concept, known to the teacher but unknown to the learner. The learner's goal is to identify Θ.

Let $\mathcal{Z}$ be a finite set, called the *teaching signal set*, consisting of the raw signals the teacher can emit. Let $\perp \notin \mathcal{Z}$ be an additional symbol representing a null observation produced when a signal cannot be parsed at the learner's current knowledge state. The learner observation set is $\mathcal{Y} = \mathcal{Z} \cup \{\perp\}$.

Definition 3.1 (Signal target map)**.** A *signal target map* is a function $\mathrm{tgt} : \mathcal{Z} \to \mathcal{C}$ that assigns to each raw teaching signal $z \in \mathcal{Z}$ the concept $\mathrm{tgt}(z) \in \mathcal{C}$ that the

signal is intended to teach. We assume that every target concept is associated with at least one raw signal, that is, $\Omega \subseteq \mathrm{im}(\mathrm{tgt})$.

For each concept $c \in \mathcal{C}$, the fiber $\mathrm{tgt}^{-1}(c) = \{z \in \mathcal{Z} : \mathrm{tgt}(z) = c\}$ is the set of all raw signals designed to teach c, representing different explanations, examples, or phrasings of the same concept. Signals in the fiber $\mathrm{tgt}^{-1}(c)$ all target the same concept and therefore have the same structural effect on the learner's acquired concept set. However, they may still differ informationally: distinct signals in the fiber can encode different information about the latent target Θ.

Remark 3.2 (Fixed signal system and notation). The raw signal alphabet $\mathcal{Z}$ and the target map tgt are treated as fixed throughout a given teaching problem. Capacity quantities introduced later therefore depend not only on the mind $\mathfrak{m}$ and the acquired concept set $\mathcal{K}$, but also on this signal system $(\mathcal{Z}, \mathrm{tgt})$. When no ambiguity arises, we suppress this dependence in the notation.

The signal target map tgt and the latent target Θ play complementary but distinct roles. The random variable $\Theta \in \Omega$ specifies what the learner must ultimately identify: the realized target concept. The map tgt specifies what each individual signal is about: a signal z with $\mathrm{tgt}(z) = c$ is designed to teach concept c, which may or may not equal Θ.

In general, signals targeting prerequisite concepts may need to be presented before signals targeting Θ itself can become usable to the learner. Thus the teacher's eventual strategy has two degrees of freedom: which concept to target, and which particular encoding of that concept to use within the fiber $\mathrm{tgt}^{-1}(c)$. Consequently, a signal may carry information about the target even when it does not directly target the concept Θ.

We now introduce the *parsing map* $\rho_{\mathfrak{m}}$, which takes a raw teaching signal together with the learner's current knowledge set and either passes the signal through, when the prerequisites are satisfied, or collapses it to the null token $\perp$ otherwise.

Definition 3.3 (Parsing map)**.** A mind $\mathfrak{m}$ is equipped with a *parsing map* $\rho_{\mathfrak{m}} : \mathcal{Z} \times 2^{\mathcal{C}} \to \mathcal{Z} \cup \{\perp\}$, where $\perp$ is a *null token* indicating that the signal is unparseable. For a signal $z \in \mathcal{Z}$ with target $c = \mathrm{tgt}(z)$ and a knowledge set $\mathcal{K} \subseteq \mathcal{C}$:

(i) $\rho_{\mathfrak{m}}(z, \mathcal{K}) = z$ if $c \in \Phi_{\mathfrak{m}}(\mathcal{K})$, equivalently, if either $c \in \mathcal{K}$ already or there exists a rule $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$ with $\mathcal{S} \subseteq \mathcal{K}$;

(ii) $\rho_{\mathfrak{m}}(z, \mathcal{K}) = \perp$ if $c \notin \Phi_{\mathfrak{m}}(\mathcal{K})$, equivalently, if $c \notin \mathcal{K}$ and no rule for c has all its prerequisites in $\mathcal{K}$.

The condition $c \in \Phi_{\mathfrak{m}}(\mathcal{K})$ is the ordered condition of Definition 2.16. A concept may have multiple prerequisite sets, and the signal is parseable if any one of them is satisfied.

Dynamics. We model teaching as a repeated interaction between a teacher and a learner unfolding over discrete rounds $t = 0, 1, 2, \ldots$. The model uses a concept-level time scale: one round represents a single instructional interaction

in which the teacher emits one raw signal, the learner observes its parsed version, and the learner's acquired concept set may be updated as a result.

We take the learner's initial acquired concept set to be the axiom set of the mind: $\mathcal{K}_0 = \mathcal{A}_{\mathfrak{m}}$. For each $t \geq 0$, the set $\mathcal{K}_t \subseteq \mathcal{C}$ denotes the concepts acquired by the learner after the first t rounds of instruction. At round $t+1$, the teacher emits a raw signal $Z_{t+1} \in \mathcal{Z}$. Given the learner's current acquired concept set $\mathcal{K}_t$, the learner observation is the parsed signal $Y_{t+1} = \rho_{\mathfrak{m}}(Z_{t+1}, \mathcal{K}_t) \in \mathcal{Y}$.

Definition 3.4 (Concept-acquisition update rule)**.** Given the parsed observation $Y_{t+1} \in \mathcal{Y}$, the learner's acquired concept set evolves according to

$$\mathcal{K}_{t+1} = \begin{cases} \mathcal{K}_t \cup \{\mathrm{tgt}(Y_{t+1})\} & \text{if } Y_{t+1} \in \mathcal{Z}, \\ \mathcal{K}_t & \text{if } Y_{t+1} = \bot. \end{cases}$$

Under this update rule, each round can add at most one newly acquired concept, namely the concept targeted by the parsed signal when parsing succeeds. Time is therefore measured in units of concept-level teaching opportunities.

The rule in Definition 3.4 has two immediate consequences. First, acquisition is monotone: $\mathcal{K}_t \subseteq \mathcal{K}_{t+1}$ for all $t \geq 0$. Second, the set $\mathcal{K}_t$ records only concepts that have been explicitly acquired through parsed instruction; it is not automatically closed under the expansion rules. Thus a concept may already be reachable from $\mathcal{K}_t$, in the sense that $c \in \Phi_{\mathfrak{m}}(\mathcal{K}_t)$, without yet belonging to $\mathcal{K}_t$ itself. The learner acquires such a concept only at a later round in which it receives a parseable signal targeting c. Therefore, $\mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ describes what is in principle reachable from the axioms, whereas the process $(\mathcal{K}_t)_{t \geq 0}$ describes what has actually been acquired over time.

Lemma 3.5 (The instructional process stays inside the reachable family)**.** *For every $t \geq 0$, one has $\mathcal{K}_t \in \mathbb{K}_{\mathfrak{m}}$ almost surely.*

Lemma 3.5 shows that the stochastic teaching process evolves within the reachable family $\mathbb{K}_{\mathfrak{m}}$. Thus the family introduced in Section 2.3 not only describes structurally feasible knowledge states but also forms the natural state space for the instructional dynamics.

Definition 3.6 (Admissible teaching strategy)**.** An *admissible teaching strategy* is a sequence of stochastic kernels

$$\kappa_{t+1}(\cdot \mid \theta, y_1, \ldots, y_t) \in \Delta(\mathcal{Z}), \qquad t \geq 0,$$

so that, conditional on the realized target $\Theta = \theta$ and the parsed history $(Y_1, \ldots, Y_t) = (y_1, \ldots, y_t)$, the teacher chooses the next raw signal Z_{t+1} according to κ_{t+1}.

Because the learner's epistemic objective is to identify the latent target Θ, it maintains at each time t a belief over the possible target concepts. This belief is updated from the parsed observations $Y_1, \ldots, Y_t$, rather than from the raw teacher

emissions $Z_1, \ldots, Z_t$, which are not directly observed by the learner. Accordingly, define the learner's information filtration by

$$\mathcal{F}_t = \sigma(Y_1, \ldots, Y_t) \subseteq \mathcal{F}, \qquad t \geq 1,$$

and set $\mathcal{F}_0 = \{\varnothing, \Omega_0\}$. Given a fixed prior π_0 and a fixed admissible teaching strategy, the learner's posterior at time t is the random probability vector $\pi_t \in \Delta(\Omega)$ defined by

$$\pi_t(c) = \mathbb{P}(\Theta = c \mid \mathcal{F}_t), \qquad c \in \Omega.$$

The conditional probability is taken with respect to the probability law induced by the prior π_0 and the admissible teaching strategy. Thus the learner is modeled as Bayesian: its belief state at time t is the posterior distribution of the latent target given the parsed observation history.

Definition 3.7 (Learning state)**.** A *learning state* at time t is a pair $(\mathcal{K}_t, \pi_t)$ where

(i) $\mathcal{K}_t \subseteq \mathcal{C}$ is the learner's acquired concept set at time t;
(ii) $\pi_t \in \Delta(\Omega)$ is the learner's posterior belief over target concepts.

Thus the learning state records both dimensions of progress in the teaching process: structural progress, captured by the acquired concept set $\mathcal{K}_t$, and epistemic progress, captured by the posterior belief π_t about the latent target. The stochastic teaching dynamics therefore evolve on the product space $\mathbb{K}_{\mathfrak{m}} \times \Delta(\Omega)$.

Definition 3.8 (Completion)**.** Teaching is *complete* at time τ if both

(i) *target acquisition*: $\Theta \in \mathcal{K}_\tau$;
(ii) *identification*: $\pi_\tau(\Theta) = 1$.

The framework therefore distinguishes three related notions. First, the understanding horizon $\mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ is the set of concepts that are in principle reachable from the axioms under the expansion rules of the mind. Second, the time-indexed set $\mathcal{K}_t$ records which concepts have actually been acquired through instruction by time t. Third, the completion condition of Definition 3.8 formalizes successful teaching of the target: it requires both acquisition of the target, $\Theta \in \mathcal{K}_\tau$, and identification of the target, $\pi_\tau(\Theta) = 1$. Acquisition without identification corresponds to having acquired a concept without yet knowing that it is the intended target. Identification without acquisition corresponds to knowing which concept is intended without yet having reached it. Completion requires both.

Example 3.9 (A full teaching interaction)**.** Let $\mathcal{C} = \{a, b, c, d\}$ with the informal readings

$$a = \text{counting}, \qquad b = \text{addition}, \qquad c = \text{arrays}, \qquad d = \text{multiplication}.$$

Fix Mind 1 from Example 2.4, with axiom set $\mathcal{A}_{\mathfrak{m}} = \{a\}$ and expansion rules

$$\{a\} \Rightarrow b, \qquad \{b\} \Rightarrow c, \qquad \{b, c\} \Rightarrow d.$$

Let $\Omega = \{b, c, d\}$, let the prior be uniform on Ω, and let the teacher use the deterministic policy

$$\Theta = b: \ (Z_1, Z_2, Z_3) = (z_b^{(1)}, z_b^{(1)}, z_b^{(1)}),$$
$$\Theta = c: \ (Z_1, Z_2, Z_3) = (z_b^{(1)}, z_c^{(1)}, z_c^{(1)}),$$
$$\Theta = d: \ (Z_1, Z_2, Z_3) = (z_b^{(1)}, z_c^{(1)}, z_d^{(1)}),$$

where $\mathrm{tgt}(z_b^{(1)}) = b$, $\mathrm{tgt}(z_c^{(1)}) = c$, and $\mathrm{tgt}(z_d^{(1)}) = d$. Suppose the realized target is $\Theta = d$. The learner starts from

$$\mathcal{K}_0 = \{a\}, \qquad \pi_0(b) = \pi_0(c) = \pi_0(d) = \frac{1}{3}.$$

At $t = 0$, the teacher emits $Z_1 = z_b^{(1)}$. Since $b \in \Phi_{\mathtt{m}}(\{a\})$, the signal is parseable, so

$$Y_1 = z_b^{(1)} \qquad \text{and} \qquad \mathcal{K}_1 = \{a, b\}.$$

Because the same first signal is prescribed under all three targets, the observation $Y_1 = z_b^{(1)}$ does not yet distinguish among them, and therefore $\pi_1 = \pi_0$.

At $t = 1$, the teacher emits $Z_2 = z_c^{(1)}$. Since b has already been acquired, the concept c is now ordered, so the signal is parseable:

$$Y_2 = z_c^{(1)} \qquad \text{and} \qquad \mathcal{K}_2 = \{a, b, c\}.$$

Under the stated policy, the history $(Y_1, Y_2) = (z_b^{(1)}, z_c^{(1)})$ is inconsistent with $\Theta = b$. Hence the posterior assigns zero mass to b and splits mass equally between c and d:

$$\pi_2(b) = 0, \qquad \pi_2(c) = \pi_2(d) = \frac{1}{2}.$$

At $t = 2$, the teacher emits $Z_3 = z_d^{(1)}$. Since both b and c are now present, the concept d is ordered, so the signal is parseable:

$$Y_3 = z_d^{(1)} \qquad \text{and} \qquad \mathcal{K}_3 = \{a, b, c, d\}.$$

Now the full observation history is consistent only with $\Theta = d$, so $\pi_3 = \delta_d$.

Thus teaching is complete at time $\tau = 3$: the learner has both acquired the target, $\Theta = d \in \mathcal{K}_3$, and identified it, $\pi_3(d) = 1$. This example illustrates the distinction between structural acquisition, encoded by the process $(\mathcal{K}_t)$, and epistemic identification, encoded by the posterior process (π_t).

3.2 The Epistemic Arrow of Time

We now formalize the epistemic component of the teaching dynamics. The key question is how the learner's uncertainty about the latent target evolves as parsed observations accumulate over time. This motivates the term *epistemic arrow of time*: although particular observations may be uninformative, posterior uncertainty can only decrease in conditional expectation under Bayesian updating. The information-theoretic notions used below are standard; see, for example, [Cover and Thomas, 2006, §2].

Information-theoretic quantities. Let X and Y be discrete random variables on a probability space $(\Omega_0, \mathcal{F}, \mathbb{P})$ taking values in finite or countable sets $\mathcal{X}$ and $\mathcal{Y}$. We adopt the convention $0 \log 0 = 0$. The Shannon entropy of X is

$$\mathbb{H}(X) = -\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) \log \mathbb{P}(X = x).$$

For a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$, define the *pathwise conditional entropy* of X given $\mathcal{G}$ by

$$\mathbb{H}(X \mid \mathcal{G}) = -\sum_{x \in \mathcal{X}} \mathbb{P}(X = x \mid \mathcal{G}) \log \mathbb{P}(X = x \mid \mathcal{G}).$$

Its expectation $\mathbb{E}[\mathbb{H}(X \mid \mathcal{G})]$ is the usual conditional entropy. For brevity, we write

$$\mathbb{H}(X \mid Y) = \mathbb{E}[\mathbb{H}(X \mid \sigma(Y))].$$

The mutual information between X and Y is

$$\mathbb{I}(X; Y) = \mathbb{H}(X) - \mathbb{H}(X \mid Y).$$

The conditional mutual information given $\mathcal{G}$ is

$$\mathbb{I}(X; Y \mid \mathcal{G}) = \mathbb{H}(X \mid \mathcal{G}) - \mathbb{E}[\mathbb{H}(X \mid \mathcal{G} \vee \sigma(Y)) \mid \mathcal{G}].$$

In the teaching model, Θ is Ω-valued and the learner filtration is

$$\mathcal{F}_t = \sigma(Y_1, \ldots, Y_t).$$

We define the *epistemic entropy* at time t by

$$H_t = \mathbb{H}(\Theta \mid \mathcal{F}_t).$$

Since $\pi_t(c) = \mathbb{P}(\Theta = c \mid \mathcal{F}_t)$, this may be written as

$$H_t = -\sum_{c \in \Omega} \pi_t(c) \log \pi_t(c) \qquad \text{a.s.}$$

Thus H_t is the Shannon entropy of the learner posterior at time t.

Proposition 3.10 (Entropy drop equals information flow)**.** *The one-round expected reduction in epistemic entropy satisfies* $\mathbb{E}[H_t - H_{t+1} \mid \mathcal{F}_t] = \mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t)$.

Proposition 3.10 expresses a conservation principle: the expected reduction in posterior uncertainty about the target is equal to the conditional mutual information conveyed by the next parsed observation. In other words, expected learning progress in one round is precisely the amount of information that Y_{t+1} carries about the target Θ.

Theorem 3.11 (Epistemic arrow of time)**.** *The epistemic entropy process* $(H_t)_{t \geq 0}$ *is a supermartingale:*

$$\mathbb{E}[H_{t+1} \mid \mathcal{F}_t] \leq H_t,$$

with equality if and only if Y_{t+1} *is independent of* Θ *given* $\mathcal{F}_t$.

Theorem 3.11 formalizes the epistemic arrow of time: posterior uncertainty decreases in conditional expectation, although along particular sample paths it may increase after a realized observation. Equality holds when the next observation carries no information about the target.

Remark 3.12 (Bayesian modeling choice). By defining $\pi_t(c) = \mathbb{P}(\Theta = c \mid \mathcal{F}_t)$, we have adopted a Bayesian learner model: the learner belief is the true conditional distribution of the target given the parsed observation history. This is not the only possible choice, but it is natural here for three reasons. First, π_t uses all information contained in the observations and nothing else, so it is determined entirely by the prior π_0 and the filtration $\mathcal{F}_t$. Second, because π_t is the conditional distribution of Θ given $\mathcal{F}_t$, the epistemic entropy H_t coincides with the conditional entropy $\mathbb{H}(\Theta \mid \mathcal{F}_t)$. This makes mutual information the natural measure of learning progress: each new observation reduces posterior uncertainty by precisely $\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t)$ in conditional expectation. Third, the completion condition $\pi_\tau(\Theta) = 1$ then has a strong interpretation: the parsed observations identify the target, rather than the learner merely arriving at the correct answer by chance.

3.3 Prerequisites and the Relativity of Randomness

The epistemic arrow of time in Theorem 3.11 describes how uncertainty evolves once observations are received. It does not, however, determine what the learner actually observes. In the teaching model the learner does not observe the raw teacher signal Z_{t+1} directly; instead it receives the parsed observation $Y_{t+1} = \rho_{\mathfrak{m}}(Z_{t+1}, \mathcal{K}_t)$, where the parsing map depends on the learner's current acquired concept set. When the targeted concept is ordered, the signal passes through unchanged; when prerequisites are missing, the parser collapses the signal to the null token $\perp$. The effective information channel from Θ to the learner is therefore state dependent. In particular, the same raw broadcast may transmit usable information to one learner while being erased for another.

The next result formalizes this phenomenon. As throughout, conditional mutual-information expressions given U_{t+1} or U_{t+1}^c are understood on the event where the relevant conditioning probability is positive, and are taken to be 0 otherwise.

Theorem 3.13 (Relativity of randomness)**.** *Let $C_{t+1} = \mathrm{tgt}(Z_{t+1})$ be the targeted concept, and define the unparseability event $U_{t+1} = \{C_{t+1} \notin \Phi_{\mathfrak{m}}(\mathcal{K}_t)\}$. Assume that on parseable rounds the raw teacher signal is informative about the latent target: $\mathbb{I}(\Theta; Z_{t+1} \mid \mathcal{F}_t, U_{t+1}^c) > 0$. Then the learner's per-round information transfer exhibits an eventwise dichotomy:*

$$\begin{aligned} &\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t, U_{t+1}) = 0 \qquad \text{(erasure)}, \\ &\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t, U_{t+1}^c) > 0 \qquad \text{(informative)}. \end{aligned}$$

Theorem 3.13 shows that the usable information in a teaching signal is state dependent. Under the parsing map $\rho_{\mathfrak{m}}$, if the targeted concept is unordered at

$\mathcal{K}_t$, then the parsed observation collapses to $\perp$; by Theorem 3.13, the learner therefore receives no usable information about the target on that event. By contrast, on parseable rounds the same raw broadcast may transmit strictly positive information. In this precise sense, the informational status of a signal is relative to the learner's structural capacity to decode it.

This relativity is consistent with classical information theory. Randomness has always been observer dependent: a ciphertext appears as pure noise without the cryptographic key [Shannon, 1949], and conditional mutual information formalizes the dependence of information on what is known [Cover and Thomas, 2006]. What is distinctive here is the mechanism that generates this dependence: the learner's decoding power is governed by the combinatorial closure operator $\Phi_{\mathfrak{m}}$, so prerequisite topology directly determines when the channel behaves as identity and when it behaves as erasure.

The notion of *mind-relative randomness* introduced earlier is related to, but distinct from, the combinatorial distinction between ordered and unordered concepts from Definition 2.16. The latter is a structural property of the targeted concept relative to the learner's acquired concept set, whereas the former is an epistemic property of the observation process relative to the latent target Θ.

In the sharp parsing model, if the teacher targets a concept $C_{t+1} = \mathrm{tgt}(Z_{t+1})$ that is unordered at the current acquired concept set, $C_{t+1} \notin \Phi_{\mathfrak{m}}(\mathcal{K}_t)$, then the parser maps every such raw signal to the same null observation: $Y_{t+1} = \perp$ almost surely on that event. Thus all distinctions among those raw signals are erased at the learner end of the channel.

However, the appearance of $\perp$ does not by itself imply mind-relative randomness. Even though the symbol $\perp$ contains no internal distinctions, the event $\{Y_{t+1} = \perp\}$ may still convey information about the target Θ. In particular, if the teacher's targeting rule depends on Θ, then the probability that the teacher selects a concept outside $\Phi_{\mathfrak{m}}(\mathcal{K}_t)$ may vary with Θ, and observing $\perp$ can update the learner's posterior belief.

Conversely, an ordered round need not be informative. If $C_{t+1} \in \Phi_{\mathfrak{m}}(\mathcal{K}_t)$, then the signal is parseable and $Y_{t+1} = Z_{t+1}$. But even in this case the parsed observation may still be mind-random if, conditional on the public history $\mathcal{F}_t$, the teacher's policy induces the same distribution of Y_{t+1} under every possible target. Equivalently, $\Theta \perp\!\!\!\perp Y_{t+1} \mid \mathcal{F}_t$. Thus parseability and informativeness are logically distinct: an unordered round may still be informative through the occurrence of erasure, while an ordered round may be uninformative if the parsed signal distribution does not depend on the target.

An immediate consequence of sharp parsing is that repeated rephrasings of the same unordered concept do not help. If $c \notin \Phi_{\mathfrak{m}}(\mathcal{K}_t)$, then every raw signal targeting c collapses to the null observation $\perp$, regardless of how many distinct encodings or phrasings are available (see Corollary B.4). Thus, on that event, repetition and rephrasing do not reduce epistemic uncertainty about the target.

Combined with the prerequisite gating established in Theorem 3.13, these observations formalize a central thesis of the framework: whether a broadcast conveys usable information is not an intrinsic property of the signal itself, but of

the interaction among the signal, the learner's current acquired concept set, and the teacher's policy.

Remark 3.14. The relativity of randomness established in Theorem 3.13 suggests a broader perspective in which randomness itself becomes observer dependent. The parsing map $\rho_{\mathfrak{m}}$ determines, for each mind and acquired state, which signals are informative and which collapse to noise. In this sense, randomness is not an intrinsic property of a signal but a relation between the signal and the observer's structure of understanding.

4 Speed Limits of Teaching

We now derive the quantitative speed limits of the teaching model. Two obstructions coexist. The first is *structural*: the learner must acquire enough prerequisite concepts for the target to become reachable. The second is *epistemic*: the learner must resolve uncertainty about which target concept the teacher intends. The purpose of this section is to formalize both obstructions and combine them into a single lower bound on the expected completion time.

Fix a mind $\mathfrak{m} = (\mathcal{C}, \mathcal{A}_{\mathfrak{m}}, \mathcal{E}_{\mathfrak{m}})$ and a finite target set $\Omega \subseteq \mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$. Thus every target under consideration lies in the learner understanding horizon.

4.1 Identification and state-dependent capacity

Recall that $\Theta : \Omega_0 \to \Omega$ is the realized target concept and that the learner observes the parsed history $\mathcal{F}_t = \sigma(Y_1, \dots, Y_t)$. The learner epistemic objective is to identify Θ from this history. We say that identification occurs at time t if $\pi_t(\Theta) = 1$, equivalently, $\mathbb{H}(\Theta \mid \mathcal{F}_t) = 0$. Since completion additionally requires target acquisition, identification is a strictly weaker requirement than full teaching completion.

Definition 4.1 (Identification stopping time)**.** A random time τ_{id} is an *identification stopping time* if:

(i) τ_{id} is an $(\mathcal{F}_t)$-stopping time;
(ii) $\mathbb{P}(\tau_{\mathrm{id}} < \infty) = 1$;
(iii) Θ is $\mathcal{F}_{\tau_{\mathrm{id}}}$-measurable.

Equivalently, $\mathbb{H}(\Theta \mid \mathcal{F}_{\tau_{\mathrm{id}}}) = 0$ almost surely.

Because the parsing map $\rho_{\mathfrak{m}}$ depends on the learner's current acquired concept set $\mathcal{K}_t$, the effective learner-side channel is state dependent. At early stages many raw signals may collapse to $\perp$, whereas later the same signals may pass through unchanged once the relevant prerequisites have been acquired.

Let $\mathbb{K}_{\mathfrak{m}}$ be the reachable family introduced in Definition 2.22. For each reachable acquired concept set $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}}$, define the ordered raw-signal set $\mathcal{Z}_{\mathrm{ord}}(\mathcal{K}) = \{z \in \mathcal{Z} : \mathrm{tgt}(z) \in \Phi_{\mathfrak{m}}(\mathcal{K})\}$. Under sharp parsing, signals in $\mathcal{Z}_{\mathrm{ord}}(\mathcal{K})$ pass through unchanged, while all other raw signals collapse to $\perp$.

This leads to the following learner-side capacity notion.

Definition 4.2 (State-dependent parsed entropy bound)**.** For each $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}}$, define

$$C_{\mathfrak{m}}(\mathcal{K}) = \sup \left\{ \mathbb{H}\Big(\rho_{\mathfrak{m}}(Z, \mathcal{K})\Big) : Z \text{ is an } \mathcal{Z}\text{-valued random variable} \right\}.$$

The parsed entropy bound $C_{\mathfrak{m}}(\mathcal{K})$ also depends on the signal system $(\mathcal{Z}, \mathrm{tgt})$. Throughout the paper this instructional interface is treated as fixed, and we therefore suppress this dependence in the notation. For a given interface, the variation of $C_{\mathfrak{m}}(\mathcal{K})$ across acquired concept sets is endogenous to the learner state, whereas its numerical level is determined jointly by the mind $\mathfrak{m}$ and the signal system $(\mathcal{Z}, \mathrm{tgt})$. Thus $C_{\mathfrak{m}}(\mathcal{K})$ should be understood as a property of the pair $(\mathfrak{m}, (\mathcal{Z}, \mathrm{tgt}))$ evaluated at state $\mathcal{K}$.

Thus $C_{\mathfrak{m}}(\mathcal{K})$ is the largest Shannon entropy that a one-round parsed observation $\rho_{\mathfrak{m}}(Z, \mathcal{K})$ can attain at the learner end of the channel when the acquired concept set is $\mathcal{K}$, as the law of the raw input signal Z ranges over all $\mathcal{Z}$-valued distributions.

Proposition 4.3 (Statewise one-round information bound)**.** *For every $t \geq 0$, $\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t) \leq C_{\mathfrak{m}}(\mathcal{K}_t)$ almost surely.*

Proposition 4.3 shows that the learner's per-round information gain about the target is bounded by the capacity $C_{\mathfrak{m}}(\mathcal{K}_t)$, which depends on the learner's acquired concept set at time t. As the learner acquires more concepts, the set of parseable signals grows, and the capacity may increase. The bound is therefore not static: structural progress expands the effective channel through which teaching occurs. This coupling between structural progress and informational capacity is the mechanism through which prerequisites govern the speed of teaching.

The next lemma shows that the learner-side channel can only improve as the learner acquires more concepts.

Lemma 4.4 (Monotonicity of the state-dependent bound)**.** *If $\mathcal{K}, \mathcal{K}' \in \mathbb{K}_{\mathfrak{m}}$ satisfy $\mathcal{K} \subseteq \mathcal{K}'$, then $C_{\mathfrak{m}}(\mathcal{K}) \leq C_{\mathfrak{m}}(\mathcal{K}')$.*

Lemma 4.4 reflects a basic property of the parsing model: acquiring additional concepts cannot reduce the learner's ability to decode signals. When the acquired concept set grows, previously parseable signals remain parseable, and additional signals may become usable. Consequently the entropy of the parsed observation, and therefore the effective channel capacity, cannot decrease as the learner acquires additional concepts.

This monotonicity admits a stronger statistical interpretation. To formalize it, we use the Blackwell order on statistical experiments [Blackwell, 1953]. Informally, one experiment Blackwell-dominates another if the latter can be obtained from the former by garbling, that is, by post-processing through a stochastic map independent of the underlying state. Equivalently, the dominating experiment is at least as informative for every statistical decision problem.

Definition 4.5 (Blackwell domination)**.** Let Ω be a finite state space, and let $\mathbb{W} : \Omega \to \Delta(\mathcal{Y})$ and $\mathbb{W}' : \Omega \to \Delta(\mathcal{Y}')$ be two statistical experiments. We say that

$\mathbb{W}$ *Blackwell-dominates* $\mathbb{W}'$ if there exists a Markov kernel $\mathbb{G} : \mathcal{Y} \to \Delta(\mathcal{Y}')$ such that for every $\omega \in \Omega$, $\mathbb{W}'(\cdot \mid \omega) = \sum_{y \in \mathcal{Y}} \mathbb{G}(\cdot \mid y)\mathbb{W}(y \mid \omega)$. Equivalently, $\mathbb{W}'$ is a garbling of $\mathbb{W}$.

Theorem 4.6 (Blackwell order on acquired concept sets)**.** *Fix $t \geq 0$ and a public history $h_t = (y_1, \ldots, y_t) \in \mathcal{Y}^t$ with $\mathbb{P}((Y_1, \ldots, Y_t) = h_t) > 0$. For each $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}}$, let $\mathbb{W}_{\mathcal{K},h_t}$ denote the statistical experiment from Θ to the parsed observation induced by the conditional raw-signal law*

$$\mathbb{P}\Big(Z_{t+1} \in \cdot \mid \Theta = \omega,\, (Y_1, \ldots, Y_t) = h_t\Big), \qquad \omega \in \Omega.$$

If $\mathcal{K} \subseteq \mathcal{K}'$, then $\mathbb{W}_{\mathcal{K}',h_t}$ Blackwell-dominates $\mathbb{W}_{\mathcal{K},h_t}$.

Theorem 4.6 holds for each realized public history h_t separately. Thus the ordering of acquired concept sets is pathwise rather than merely averaged: conditional on any history for which the next-round raw-signal law is defined, the parsed experiment induced by a larger acquired concept set Blackwell-dominates the parsed experiment induced by a smaller one.

This theorem strengthens Lemma 4.4. The monotonicity of $C_{\mathfrak{m}}(\mathcal{K})$ says that larger acquired concept sets permit weakly greater parsed entropy. Theorem 4.6 shows more: they induce uniformly more informative experiments in the sense of statistical decision theory. The universal-broadcast theorem of Theorem 5.6 will show that this dependence on the learner prerequisite structure cannot, in general, be eliminated by a common broadcast curriculum.

4.2 Structural and epistemic lower bounds

We now combine the structural and epistemic constraints of the model to derive a single lower bound on teaching time.

Definition 4.7 (Structural distance to a target concept)**.** For $c \in \mathcal{U}_{\mathfrak{m}}$, define

$$L_{\mathfrak{m}}(c) = \min\Big\{ L \geq 0 : \exists\, \mathcal{K}_0, \ldots, \mathcal{K}_L \in \mathbb{K}_{\mathfrak{m}},\ u_0, \ldots, u_{L-1} \in \mathcal{U}_{\mathfrak{m}} \text{ such that}$$

$$\mathcal{K}_0 = \mathcal{A}_{\mathfrak{m}},\ c \in \mathcal{K}_L,\ \mathcal{K}_{i+1} = \mathcal{K}_i \cup \{u_i\},\ u_i \in \Phi_{\mathfrak{m}}(\mathcal{K}_i) \setminus \mathcal{K}_i, \quad i = 0, \ldots, L-1 \Big\}.$$

The quantity $L_{\mathfrak{m}}(c)$ measures the shortest prerequisite-respecting route from the axioms to a state containing c. It therefore gives the natural structural benchmark against which any completion time must be compared. The first fundamental constraint on teaching time is structural: the learner must traverse the prerequisite chain before the target can be acquired.

Proposition 4.8 (Structural barrier)**.** *Let τ be any completion time in the sense of Definition 3.8. Then $\tau \geq L_{\mathfrak{m}}(\Theta)$ almost surely. Consequently, $\mathbb{E}[\tau] \geq \mathbb{E}[L_{\mathfrak{m}}(\Theta)]$.*

Proposition 4.8 is the purely geometric obstruction in the model. Regardless of how informative the signals are, the learner cannot complete teaching before traversing a prerequisite-respecting path to a state containing the realized target. Since each round adds at most one concept, the shortest such path gives an unavoidable lower bound on completion time.

To control the epistemic obstruction, we next aggregate the information gained across all rounds up to identification τ_{id}. The point is that the per-round entropy drop identity from Proposition 3.10 telescopes over time.

Lemma 4.9 (Total information required for identification)**.** *Let τ_{id} be an identification stopping time. Then*

$$\mathbb{E}\left[\sum_{t=0}^{\tau_{\mathrm{id}}-1} \mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t)\right] = \mathbb{H}(\Theta).$$

Lemma 4.9 says that identification must pay for the full initial uncertainty of the target. The cumulative conditional mutual information transmitted through the parsed observations up to identification is the entropy of Θ. Thus the learner cannot identify the target until enough usable information has flowed through the learner-side channel to resolve all initial uncertainty.

The next step is to combine this accounting identity with the statewise capacity bound from Proposition 4.3. This converts total required information into a lower bound expressed in terms of the learner trajectory through the reachable family.

Proposition 4.10 (Trajectory information budget)**.** *Let τ_{id} be any identification stopping time. Then*

$$\mathbb{H}(\Theta) \le \mathbb{E}\left[\sum_{t=0}^{\tau_{\mathrm{id}}-1} C_{\mathfrak{m}}(\mathcal{K}_t)\right].$$

Proposition 4.10 is the dynamic information budget of the model. The total target uncertainty cannot exceed the cumulative parsed capacity along the states visited before identification. In this sense, a curriculum may need to spend rounds building the decoder before it can effectively use it: the states through which the learner passes determine the rate at which target information can be transmitted.

We define $C_{\mathfrak{m}}^{\max} = \max_{\mathcal{K}\in\mathbb{K}_{\mathfrak{m}}} C_{\mathfrak{m}}(\mathcal{K})$. This maximum is well defined because, under Assumption 2.23, the reachable family $\mathbb{K}_{\mathfrak{m}}$ is finite by Proposition 2.24.

Theorem 4.11 (Global structural-information lower bound)**.** *Let τ be any completion time, then*

$$\mathbb{E}[\tau] \ge \max\left\{\mathbb{E}[L_{\mathfrak{m}}(\Theta)], \frac{\mathbb{H}(\Theta)}{C_{\mathfrak{m}}^{\max}}\right\}.$$

Theorem 4.11 is the central speed law of the framework. Teaching is constrained simultaneously by prerequisite geometry and by information transmission. The lower bound takes the form of a maximum rather than an additive sum because structural progress may itself convey information about the target. Nevertheless both bottlenecks must be cleared.

Assumption 4.12 (Structural signal availability)**.** *For every concept $u \in \mathcal{U}_{\mathfrak{m}}$, there exists a raw signal $z \in \mathcal{Z}$ such that* $\mathrm{tgt}(z) = u$.

Earlier we required only that every possible target concept admit a corresponding signal, that is, $\Omega \subseteq \mathrm{im}(\mathrm{tgt})$. Assumption 4.12 is stronger: it requires signals for all concepts in the understanding horizon $\mathcal{U}_{\mathfrak{m}}$, including intermediate prerequisites. This assumption ensures that the teacher can implement any valid ordered curriculum by emitting signals targeting the concepts that must be acquired along the path to the target.

Proposition 4.13 (Direct target signaling collapses the epistemic term in the baseline model)**.** *Let $\Omega_+ = \{c \in \Omega : \pi_0(c) > 0\}$ be the support of the prior.*

(i) *If $\Omega \subseteq \mathrm{im}(\mathrm{tgt})$, then $\frac{\mathbb{H}(\Theta)}{C_{\mathfrak{m}}^{\max}} \leq 1$.*

(ii) *Under Assumption 4.12, there exists an admissible teaching strategy with completion time τ satisfying $\tau \leq L_{\mathfrak{m}}(\Theta) + 1$ almost surely, and hence $\mathbb{E}[\tau] \leq \mathbb{E}[L_{\mathfrak{m}}(\Theta)] + 1$.*

Proposition 4.13 clarifies the role of the information-theoretic layer in the baseline model. Once the learner has structurally reached the realized target, a single target-specific signal suffices for identification. Thus the dominant obstruction is typically structural: the learner must first acquire the prerequisites that make the target concept reachable.

The information-theoretic analysis nevertheless remains essential. It explains why target-specific instruction is ineffective before the relevant prerequisites are in place, and it provides a principled way to compare the informativeness of different acquired states through Blackwell dominance. In this view, structural progress builds the decoder, and information transmission becomes effective only after that decoder exists.

Example 4.14 (A common prerequisite can open a parseable identification channel)**.** Consider the mind $\mathfrak{m} = (\mathcal{C}, \mathcal{A}_{\mathfrak{m}}, \mathcal{E}_{\mathfrak{m}})$ with

$$\mathcal{C} = \{a, b, d_1, d_2, d_3, d_4\}, \qquad \mathcal{A}_{\mathfrak{m}} = \{a\},$$

and expansion rules $\{a\} \Rightarrow b$, $\{b\} \Rightarrow d_j$, $j = 1, 2, 3, 4$. Thus b is a common prerequisite, and once b has been acquired, any of the four target concepts d_1, d_2, d_3, d_4 becomes reachable in one additional step. Let $\Omega = \{d_1, d_2, d_3, d_4\}$ with the uniform prior. Then $\mathbb{H}(\Theta) = \log 4 = 2$. For each $j = 1, 2, 3, 4$, $L_{\mathfrak{m}}(d_j) = 2$, and therefore $\mathbb{E}[L_{\mathfrak{m}}(\Theta)] = 2$. Let the raw signal alphabet be $\mathcal{Z} = \{z_b, z_1, z_2, z_3, z_4\}$, with $\mathrm{tgt}(z_b) = b$, $\mathrm{tgt}(z_j) = d_j$, $j = 1, 2, 3, 4$.

At the initial acquired concept set $\{a\}$, only b is ordered. Hence the parsed observation range is $\{z_b, \perp\}$, so $C_{\mathfrak{m}}(\{a\}) = \log 2$. At the acquired concept set $\{a, b\}$, all five raw signals are parseable, so the parsed observation range is $\{z_b, z_1, z_2, z_3, z_4\}$, and therefore $C_{\mathfrak{m}}(\{a, b\}) = \log 5$. Thus acquiring the single prerequisite b enlarges the learner effective channel from 1 bit to $\log 5$ bits per round.

Now suppose the teacher tries to identify the target immediately by sending

$$Z_1 = z_j \qquad \text{when } \Theta = d_j.$$

At the raw-signal level this would reveal the target perfectly. But at the learner initial acquired concept set $\{a\}$, none of the targets d_j is ordered, so

$$Y_1 = \rho_{\mathfrak{m}}(Z_1, \{a\}) = \bot \qquad \text{almost surely.}$$

Hence $\mathbb{I}(\Theta; Y_1 \mid \mathcal{F}_0) = 0$. Before the common prerequisite b is taught, target-specific instruction is pure erasure.

Consider instead the two-round strategy

$$Z_1 = z_b \qquad \text{for every realization of } \Theta,$$

and

$$Z_2 = z_j \qquad \text{if } \Theta = d_j.$$

After round 1, the learner has acquired the prerequisite: $\mathcal{K}_1 = \{a, b\}$. The posterior does not change, because the first signal is independent of Θ. At round 2, the signal z_j is parseable, so the learner observes $Y_2 = z_j$, acquires d_j, and identifies the target exactly. Thus $\tau = 2$ almost surely.

The lower bound of Theorem 4.11 is therefore tight in this example. Since $C_{\mathfrak{m}}^{\max} = \log 5$, one obtains

$$\mathbb{E}[\tau] \geq \max\left\{\mathbb{E}[L_{\mathfrak{m}}(\Theta)], \frac{\mathbb{H}(\Theta)}{C_{\mathfrak{m}}^{\max}}\right\} = \max\left\{2, \frac{2}{\log 5}\right\} = 2,$$

and the strategy above attains equality.

This example shows that an optimal teacher may rationally spend an entire round on structural preparation rather than on target-specific signaling, because target-specific signals are useless before the common prerequisite b has been acquired. In the baseline model, the information-theoretic term is not the binding lower bound here, since Proposition 4.13 implies that identification costs at most one additional round once the target is structurally reachable. The example nevertheless illustrates the central mechanism of the section: usable information is state dependent, and teaching may need to enlarge the learner parsed alphabet before target information can flow.

5 Structural Limits on Teaching

This section develops two consequences of the structural view of teaching. First, for a fixed learner mind, the prerequisite geometry creates threshold effects in finite-horizon teaching: below a critical time budget, completion is impossible for every strategy, while beyond that threshold success becomes feasible and, under mild assumptions, eventually likely. Second, for heterogeneous learners, structural incompatibilities generate an intrinsic inefficiency of universal broadcast

curricula: a single common sequence of signals may be forced to pay separately for prerequisites that personalized teaching would handle individually.

Taken together, these results show that the limits of teaching are not merely informational. They are already encoded in the combinatorial structure of the learner prerequisite system. That structure determines both when teaching can begin to succeed and how costly common instruction becomes across different minds.

5.1 Structural thresholds in teaching

A central question in any teaching problem is: *given a fixed time budget, what is the probability that teaching succeeds?* The prerequisite structure of the learner determines the answer. Below a certain threshold, completion is impossible for every teaching strategy. Once the time horizon exceeds that threshold, completion is no longer ruled out a priori, and under mild assumptions the optimal fixed-horizon success probability converges to one as the horizon grows.

This vanishing completion probability is not an approximation but a direct consequence of the structural barrier. It also has an immediate implication for resource allocation: when training budgets are scarce, distributing time evenly across learners may produce no completed learners at all, whereas concentrating the same budget on fewer learners can yield strictly positive output.

Recall the stochastic teaching model from Section 3.1. The target concept Θ is drawn from a prior π_0 on Ω, known to both teacher and learner. By Definition 3.8, teaching is complete at the random time τ if both

(i) the learner has acquired the target concept, $\Theta \in \mathcal{K}_\tau$;
(ii) the learner has identified the target, $\pi_\tau(\Theta) = 1$.

We therefore ask: if the teacher is given a budget of t rounds, what is the maximal probability of completing teaching within that budget? Define

$$\mathbb{V}(t) = \sup_{\text{admissible teaching strategies}} \mathbb{P}(\tau \le t).$$

Thus $\mathbb{V}(t)$ is the optimal success probability achievable with a time budget of t rounds, computed under the prior on Θ.

Recall also that for each target $c \in \Omega$, the quantity $L_{\mathfrak{m}}(c)$ denotes the structural distance from the axiom set $\mathcal{A}_{\mathfrak{m}}$ to a reachable acquired concept set containing c. Define

$$L_{\min} = \min\{L_{\mathfrak{m}}(c) : \pi_0(c) > 0\}.$$

This is the smallest structural distance among targets that can arise under the prior.

For expected completion time, the baseline model also admits the upper bound $\mathbb{E}[\tau] \le \mathbb{E}[L_{\mathfrak{m}}(\Theta)] + 1$ under Assumption 4.12 (Proposition 4.13). The fixed-horizon analysis below complements that statement by describing the threshold structure of success probabilities as a function of the time budget.

Proposition 5.1 (Zero completion below the structural threshold)**.** *For every* $t \in \mathbb{N}$*,*

$$\mathbb{V}(t) \leq \mathbb{P}\Big(L_{\mathfrak{m}}(\Theta) \leq t\Big).$$

In particular, $\mathbb{V}(t) = 0$ *for all* $t < L_{\min}$*.*

Proposition 5.1 shows that if the time budget is shorter than the structural depth of every possible target, then completion is impossible. No teaching strategy can circumvent this obstruction, because the learner cannot be moved to a reachable acquired concept set containing the realized target in so few rounds.

At the opposite extreme, if some admissible strategy completes teaching in finite expected time, then the optimal fixed-horizon success probability converges to one as the horizon grows.

Proposition 5.2 (Eventual success)**.** *If there exists an admissible teaching strategy such that* $\mathbb{E}[\tau] < \infty$*, then* $\mathbb{V}(t) \to 1$ *as* $t \to \infty$*. More concretely, for any such strategy,*

$$\mathbb{V}(t) \geq 1 - \frac{\mathbb{E}[\tau]}{t} \qquad \text{for all } t \geq 1.$$

Together, Propositions 5.1 and 5.2 describe the qualitative shape of the fixed-horizon success function $\mathbb{V}(t)$: an initial region of structural impossibility, followed by a region in which success becomes increasingly likely as the time budget grows.

To make the allocation implications transparent, it is useful to consider the special case of a deterministic target.

Let $g \in \mathcal{U}_{\mathfrak{m}}$ be fixed, and suppose that $\Theta = g$ almost surely. Then the prior is degenerate, so $\pi_t = \delta_g$ for all t. Hence identification is automatic, and completion reduces to target acquisition alone.

Define the target-acquisition time of g by $\tau_g = \inf\{t \geq 0 : g \in \mathcal{K}_t\}$, and define the fixed-horizon acquisition probability by

$$\mathbb{V}_g(t) = \sup_{\text{admissible teaching strategies}} \mathbb{P}(\tau_g \leq t).$$

Proposition 5.3 (Step function for deterministic targets)**.** *Assume that the parsing map is given by Definition 3.3 and that Assumption 4.12 holds. Then for every deterministic target* $g \in \mathcal{U}_{\mathfrak{m}}$*,*

$$\mathbb{V}_g(t) = \begin{cases} 0, & \text{if} t < L_{\mathfrak{m}}(g), \\ 1, & \text{otherwise.} \end{cases}$$

Thus, for a deterministic target, the fixed-horizon acquisition probability is a step function at the structural distance $L_{\mathfrak{m}}(g)$. Below that threshold acquisition is impossible; at and above it, acquisition can be achieved with certainty.

Remark 5.4 (Threshold versus smooth training technologies). The threshold structure above contrasts with benchmark models of human-capital accumulation in which training is represented by a smooth production technology for human capital (*e.g.*, [Ben-Porath, 1967, Becker, 1964]). In such models every marginal

unit of investment yields a positive, though possibly diminishing, return. In the present framework, prerequisite-gated learning induces a threshold technology: a teaching signal has no effect until the learner prerequisite structure admits the target concept, after which additional signals become productive. The induced production technology is therefore *non*-concave.

This threshold structure has direct implications for the allocation of training resources. Consider a decision maker who must allocate a fixed instructional budget across learners, for example a firm training workers in a specific skill or an instructor allocating tutoring hours across students. The planner has a total budget of B instructional rounds and must decide how to distribute them across N learners.

Proposition 5.5 (Allocation under structural thresholds)**.** *Assume that the parsing map is given by Definition 3.3 and that Assumption 4.12 holds. Fix a deterministic target $g \in \mathfrak{U}_{\mathfrak{m}}$ with $L = L_{\mathfrak{m}}(g) \geq 1$. Consider N identical learners and a total budget of $B \in \mathbb{N}$ instructional rounds.*

(i) *Any allocation that gives every learner fewer than L rounds yields zero completed learners.*

(ii) *There exists an allocation that gives L rounds to $\lfloor B/L \rfloor$ learners and 0 rounds to the remaining learners, and under this allocation $\lfloor B/L \rfloor$ learners complete.*

In particular, if $B < NL$ and the budget is spread so that every learner receives fewer than L rounds, then total output is zero, whereas the concentrated allocation in (ii) *yields strictly positive output whenever $B \geq L$.*

Proposition 5.5 shows that evenly spreading a fixed training budget can waste the entire budget when every learner remains below the structural threshold. By contrast, concentrating the same budget on fewer learners allows those learners to cross the threshold and produce strictly positive output. The source of this effect is structural: for a deterministic target, additional training time has no effect until the prerequisite threshold $L_{\mathfrak{m}}(g)$ is reached, at which point completion becomes possible. The zero-output region is therefore not imposed from outside the model but is a direct consequence of the learner prerequisite geometry.

For random targets, the step-function structure need not persist, because different targets may have different structural depths. What remains is the zero-completion phenomenon from Proposition 5.1: if every learner receives fewer than $L_{\min} = \min\{L_{\mathfrak{m}}(c) : \pi_0(c) > 0\}$, rounds, then the completion probability is zero regardless of the teaching strategy. The qualitative allocation lesson therefore extends beyond the deterministic case: if the available budget is spread so thinly that every learner remains below the relevant structural threshold, no learner completes.

5.2 Limits of universal broadcast curricula

The preceding subsection concerned a single learner mind. We now turn to *heterogeneous* learners whose prerequisite structures differ. In that setting, a

teacher restricted to a single broadcast curriculum cannot adapt instruction to individual minds. The next theorem shows that this restriction carries a structural cost: even when each learner can be taught efficiently by a personalized curriculum, any common broadcast may be forced to pay a linear penalty in the number of learner types.

Theorem 5.6 (Linear broadcast penalty for incompatible minds)**.** *Fix integers $k \geq 2$ and $L \geq 2$. Then one can construct*

- *a finite concept space $\mathcal{C}$,*
- *a common axiom set $\mathcal{A} \subseteq \mathcal{C}$,*
- *a finite raw-signal alphabet $\mathcal{Z}$ together with a signal target map $\mathrm{tgt} : \mathcal{Z} \to \mathcal{C}$,*
- *minds $\mathfrak{m}_1, \ldots, \mathfrak{m}_k$ on $\mathcal{C}$ with common axiom set $\mathcal{A}_{\mathfrak{m}_i} = \mathcal{A}$, $i = 1, \ldots, k$, but pairwise distinct rule sets $\mathcal{E}_{\mathfrak{m}_i}$,*
- *and a common deterministic target concept $g \in \mathcal{C}$,*

such that:

(i) *for each $i \in \{1, \ldots, k\}$, there exists a valid ordered curriculum for $\mathfrak{m}_i$ of length L whose final acquired concept set contains g;*

(ii) *if a common broadcast sequence $\Gamma = (z_1, \ldots, z_T) \in \mathcal{Z}^T$ is presented to all k minds, and if the induced acquired concept processes start from $\mathcal{K}_0^{(i)} = \mathcal{A} \quad (i = 1, \ldots, k)$, and evolve according to*

$$\mathcal{K}_{t+1}^{(i)} = \begin{cases} \mathcal{K}_t^{(i)} \cup \{\mathrm{tgt}(z_{t+1})\}, & \text{if } \mathrm{tgt}(z_{t+1}) \in \Phi_{\mathfrak{m}_i}(\mathcal{K}_t^{(i)}), \\ \mathcal{K}_t^{(i)}, & \text{otherwise,} \end{cases} \qquad t = 0, \ldots, T-1,$$

then the condition $g \in \mathcal{K}_T^{(i)}$ for every $i = 1, \ldots, k$ implies $T \geq k(L-1)+1$;

(iii) *There exists a common broadcast sequence of length $k(L-1)+1$ for which $g \in \mathcal{K}_{k(L-1)+1}^{(i)}$ for every $i = 1, \ldots, k$.*

Theorem 5.6 is an existence result. For any prescribed number k of learner types and any prescribed personalized teaching length L, one can construct k minds sharing the same axiom set and the same deterministic target concept g, but having different prerequisite structures. For each mind, the target can be acquired in L personalized rounds. However, every common broadcast sequence that succeeds for all minds must have length at least $k(L-1)+1$.

The source of the penalty is purely structural. Each mind possesses a private prerequisite chain leading to the target concept, and signals that advance one mind along its chain are unparseable for the others. Consequently a universal broadcast cannot reuse prerequisite rounds across learner types; it must effectively pay for each private chain separately. This is what generates the linear dependence of the required broadcast length on the number of learner types.

Acknowledgments. The author thanks Teng Andrea Xu for helpful discussions.

Appendix

This document contains supplementary material for the paper *A Mathematical Theory of Understanding*.

1 Introduction **1**

2 Understanding as a Closure System **4**
2.1 Derivations and Equivalence 9
2.2 Ordered and Unordered Information 11
2.3 Reachable acquired concept sets 14

3 Teaching and Learning Dynamics **16**
3.1 A Stochastic Model of Teaching 17
3.2 The Epistemic Arrow of Time 21
3.3 Prerequisites and the Relativity of Randomness 23

4 Speed Limits of Teaching **25**
4.1 Identification and state-dependent capacity 25
4.2 Structural and epistemic lower bounds 27

5 Structural Limits on Teaching **30**
5.1 Structural thresholds in teaching 31
5.2 Limits of universal broadcast curricula 33

A Proofs **35**
A.1 Proofs for Section 2 36
A.2 Proofs for Section 3 41
A.3 Proofs for Section 4 42
A.4 Proofs for Section 5 46

B Additional results **49**

A Proofs

This appendix collects the proofs omitted from the main text, organized by the section in which the corresponding result appears. Supplementary lemmas are used in the proofs but not stated in the main text are included where they arise.

A.1 Proofs for Section 2

Proof of Lemma 2.7

Let $c \in \Phi_{\mathfrak{m}}(\mathcal{K})$. If $c \in \mathcal{K}$, then $c \in \mathcal{K}' \subseteq \Phi_{\mathfrak{m}}(\mathcal{K}')$. Otherwise there exists $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$ with $\mathcal{S} \subseteq \mathcal{K}$. Since $\mathcal{K} \subseteq \mathcal{K}'$, one also has $\mathcal{S} \subseteq \mathcal{K}'$, hence $c \in \Phi_{\mathfrak{m}}(\mathcal{K}')$. □

Lemma A.1 (Directed-union continuity)**.** *If $(\mathcal{K}_\alpha)_{\alpha \in \mathcal{A}}$ is a nonempty directed family, then $\Phi_{\mathfrak{m}}\left(\bigcup_{\alpha \in \mathcal{A}} \mathcal{K}_\alpha\right) = \bigcup_{\alpha \in \mathcal{A}} \Phi_{\mathfrak{m}}(\mathcal{K}_\alpha)$.*

Proof. The inclusion $\supseteq$ follows from monotonicity of $\Phi_{\mathfrak{m}}$.

For the reverse inclusion, let $c \in \Phi_{\mathfrak{m}}(\bigcup_\alpha \mathcal{K}_\alpha)$. If $c \in \bigcup_\alpha \mathcal{K}_\alpha$, then $c \in \Phi_{\mathfrak{m}}(\mathcal{K}_\alpha)$ for some $\alpha \in \mathcal{A}$.

Otherwise there exists a finite set $\mathcal{S} \subseteq \bigcup_\alpha \mathcal{K}_\alpha$ with $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$. For each $s \in \mathcal{S}$, choose α_s such that $s \in \mathcal{K}_{\alpha_s}$. Because the family is directed and $\mathcal{S}$ is finite, there exists γ with $\bigcup_{s \in \mathcal{S}} \mathcal{K}_{\alpha_s} \subseteq \mathcal{K}_\gamma$. Hence $\mathcal{S} \subseteq \mathcal{K}_\gamma$, so $c \in \Phi_{\mathfrak{m}}(\mathcal{K}_\gamma)$. □

Proof of Proposition 2.10

(i) We first show that the collection of fixed points of $\Phi_{\mathfrak{m}}$ containing $\mathcal{K}$ is non-empty. Consider the entire concept space $\mathcal{C}$. By extensiveness, $\mathcal{C} \subseteq \Phi_{\mathfrak{m}}(\mathcal{C})$. For the reverse inclusion: every expansion rule $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$ has $c \in \mathcal{C}$ by definition, so $\Phi_{\mathfrak{m}}(\mathcal{C}) \subseteq \mathcal{C}$. Together, $\Phi_{\mathfrak{m}}(\mathcal{C}) = \mathcal{C}$, and clearly $\mathcal{K} \subseteq \mathcal{C}$. So $\mathcal{C}$ is a fixed point containing $\mathcal{K}$.

By Definition 2.8, $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}) = \bigcap \{\mathcal{F} \subseteq \mathcal{C} : \mathcal{K} \subseteq \mathcal{F}, \Phi_{\mathfrak{m}}(\mathcal{F}) = \mathcal{F}\}$. The intersection is over a non-empty collection, so $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$ is well-defined. We now show it is itself a fixed point. For any fixed point $\mathcal{F} \supseteq \mathcal{K}$ in the collection, $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}) \subseteq \mathcal{F}$, so monotonicity gives $\Phi_{\mathfrak{m}}(\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})) \subseteq \Phi_{\mathfrak{m}}(\mathcal{F}) = \mathcal{F}$. Since this holds for every such $\mathcal{F}$, we get $\Phi_{\mathfrak{m}}(\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})) \subseteq \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$. Extensiveness gives the other direction: $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}) \subseteq \Phi_{\mathfrak{m}}(\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}))$. Together: $\Phi_{\mathfrak{m}}(\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})) = \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$.

(ii) By extensiveness, the sequence $\mathcal{K} \subseteq \Phi_{\mathfrak{m}}(\mathcal{K}) \subseteq \Phi^2_{\mathfrak{m}}(\mathcal{K}) \subseteq \cdots$ is non-decreasing. Let $\mathcal{L} = \cup_{n=0}^{\infty} \Phi^n_{\mathfrak{m}}(\mathcal{K})$. By Lemma A.1:

$$\Phi_{\mathfrak{m}}(\mathcal{L}) = \Phi_{\mathfrak{m}}\left(\bigcup_{n=0}^{\infty} \Phi^n_{\mathfrak{m}}(\mathcal{K})\right) = \bigcup_{n=0}^{\infty} \Phi^{n+1}_{\mathfrak{m}}(\mathcal{K}) = \mathcal{L}$$

so $\mathcal{L}$ is a fixed point of $\Phi_{\mathfrak{m}}$ containing $\mathcal{K}$. Since $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$ is the least such fixed point, $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}) \subseteq \mathcal{L}$. Conversely, we show by induction that $\Phi^n_{\mathfrak{m}}(\mathcal{K}) \subseteq \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$ for all $n \geq 0$. Base case: $\Phi^0_{\mathfrak{m}}(\mathcal{K}) = \mathcal{K} \subseteq \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$ by definition. Inductive step: suppose $\Phi^n_{\mathfrak{m}}(\mathcal{K}) \subseteq \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$. Since $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$ is a fixed point, $\Phi_{\mathfrak{m}}(\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})) = \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$. Monotonicity then gives $\Phi^{n+1}_{\mathfrak{m}}(\mathcal{K}) = \Phi_{\mathfrak{m}}(\Phi^n_{\mathfrak{m}}(\mathcal{K})) \subseteq \Phi_{\mathfrak{m}}(\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})) = \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$. Since $\Phi^n_{\mathfrak{m}}(\mathcal{K}) \subseteq \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$ for every n, we get $\mathcal{L} = \cup_{n=0}^{\infty} \Phi^n_{\mathfrak{m}}(\mathcal{K}) \subseteq \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$.

(iii) If $\mathcal{C}$ is finite, the chain $\mathcal{K} \subseteq \Phi^1_{\mathfrak{m}}(\mathcal{K}) \subseteq \Phi^2_{\mathfrak{m}}(\mathcal{K}) \subseteq \cdots$ is an increasing sequence of subsets of $\mathcal{C}$. Whenever $\Phi^{n+1}_{\mathfrak{m}}(\mathcal{K}) \neq \Phi^n_{\mathfrak{m}}(\mathcal{K})$, the inclusion is strict, so $|\Phi^{n+1}_{\mathfrak{m}}(\mathcal{K})| \geq |\Phi^n_{\mathfrak{m}}(\mathcal{K})| + 1$. Since each set has at most $|\mathcal{C}|$ elements, strict growth can occur at most $|\mathcal{C}| - |\mathcal{K}| \leq |\mathcal{C}|$ times. Therefore $\Phi^N_{\mathfrak{m}}(\mathcal{K}) = \Phi^{N+1}_{\mathfrak{m}}(\mathcal{K})$ for some $N \leq |\mathcal{C}|$, and the chain stabilizes: $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}) = \Phi^N_{\mathfrak{m}}(\mathcal{K})$. □

Proof of Proposition 2.11

We first show that $\mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ satisfies (i) to (iii). (i) By Proposition 2.10(ii), $\mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}}) = \bigcup_{n\geq 0} \Phi_{\mathfrak{m}}^n(\mathcal{A}_{\mathfrak{m}})$. Since $\Phi_{\mathfrak{m}}^0(\mathcal{A}_{\mathfrak{m}}) = \mathcal{A}_{\mathfrak{m}}$, it follows that $\mathcal{A}_{\mathfrak{m}} \subseteq \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$. (ii) Let $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$ and suppose $\mathcal{S} \subseteq \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$. By Proposition 2.10(i), $\mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ is a fixed point of $\Phi_{\mathfrak{m}}$, so $\Phi_{\mathfrak{m}}(\mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})) = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$. Since $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$ and $\mathcal{S} \subseteq \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$, the definition of $\Phi_{\mathfrak{m}}$ gives $c \in \Phi_{\mathfrak{m}}(\mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})) = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$. (iii) Let $\mathcal{F} \subseteq \mathcal{C}$ satisfy (i) and (ii). Then $\mathcal{A}_{\mathfrak{m}} \subseteq \mathcal{F}$. Moreover, if $c \in \Phi_{\mathfrak{m}}(\mathcal{F})$, then either $c \in \mathcal{F}$, or there exists $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$ with $\mathcal{S} \subseteq \mathcal{F}$, in which case (ii) gives $c \in \mathcal{F}$. Hence $\Phi_{\mathfrak{m}}(\mathcal{F}) \subseteq \mathcal{F}$. By extensiveness of $\Phi_{\mathfrak{m}}$, we also have $\mathcal{F} \subseteq \Phi_{\mathfrak{m}}(\mathcal{F})$. Therefore $\Phi_{\mathfrak{m}}(\mathcal{F}) = \mathcal{F}$, so $\mathcal{F}$ is a fixed point of $\Phi_{\mathfrak{m}}$ containing $\mathcal{A}_{\mathfrak{m}}$. Since $\mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ is the intersection of all such fixed points, we conclude that $\mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}}) \subseteq \mathcal{F}$. Thus $\mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ is the smallest set satisfying (i) and (ii).

Finally, suppose $\mathcal{U}$ and $\mathcal{U}'$ both satisfy (i)-(iii). Since $\mathcal{U}'$ satisfies (i) and (ii), the minimality property (iii) for $\mathcal{U}$ implies $\mathcal{U} \subseteq \mathcal{U}'$. By symmetry, $\mathcal{U}' \subseteq \mathcal{U}$. Hence $\mathcal{U} = \mathcal{U}'$. □

Proof of Theorem 2.14

The proof of Theorem 2.14 relies on the finiteness of derivation trees, which we establish first.

Lemma A.2 (Finiteness of derivations)**.** *Every derivation tree in the sense of Definition 2.12 is finite.*

Proof of Lemma A.2. Assume for contradiction that the derivation tree is infinite. By Definition 2.12(i), every node has finitely many children, since each prerequisite set is finite. Thus the tree is finitely branching. By König's lemma [Diestel, 2024, Lemma 8.1.2], every infinite finitely branching tree has an infinite descending path. This contradicts the well-foundedness requirement in Definition 2.12. Therefore the tree is finite. □

Proof of Theorem 2.14. For ($\Leftarrow$), suppose $\mathcal{K} \vdash_{\mathfrak{m}} c$. By Lemma A.2, the derivation tree is finite. We argue by induction on its height. If the height is 0, then either $c \in \mathcal{K}$, or $(\varnothing, c) \in \mathcal{E}_{\mathfrak{m}}$; in either case $c \in \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$. For the induction step, if the root uses a rule $(\mathcal{S}, c)$ and each child label $s \in \mathcal{S}$ has a derivation of smaller height, then by the induction hypothesis $\mathcal{S} \subseteq \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$, hence $c \in \Phi_{\mathfrak{m}}(\mathrm{cl}_{\mathfrak{m}}(\mathcal{K})) = \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$.

For ($\Rightarrow$), let $\mathcal{D} = \{d \in \mathcal{C} : \mathcal{K} \vdash_{\mathfrak{m}} d\}$. We show that $\mathcal{D}$ is a fixed point of $\Phi_{\mathfrak{m}}$ containing $\mathcal{K}$. First, $\mathcal{K} \subseteq \mathcal{D}$: for any $c \in \mathcal{K}$, the single-node tree with root labeled c is a valid derivation, so $c \in \mathcal{D}$. Next, we will show that $\Phi_{\mathfrak{m}}(\mathcal{D}) \subseteq \mathcal{D}$. Let $c \in \Phi_{\mathfrak{m}}(\mathcal{D})$. If $c \notin \mathcal{D}$, then there exists $(\mathcal{S}, c) \in \mathcal{E}_{\mathfrak{m}}$ with $\mathcal{S} \subseteq \mathcal{D}$. For each $s \in \mathcal{S}$, choose a derivation of s from $\mathcal{K}$ and attach them below a new root labeled c. This gives a derivation of c, contradiction. If $\mathcal{S} = \varnothing$, the new root has no children and still forms a valid derivation. Thus, $c \in \mathcal{D}$, and $\Phi_{\mathfrak{m}}(\mathcal{D}) \subseteq \mathcal{D}$. By extensiveness, $\mathcal{D} \subseteq \Phi_{\mathfrak{m}}(\mathcal{D})$, so $\mathcal{D}$ is a fixed point containing $\mathcal{K}$. Therefore $\mathrm{cl}_{\mathfrak{m}}(\mathcal{K}) \subseteq \mathcal{D}$. By definition of $\mathcal{D}$, this means that if $c \in \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$, then $\mathcal{K} \vdash_{\mathfrak{m}} c$. □

Proof of Theorem 2.15

We first recall the abstract definition of algebraic closure operator.

Definition A.3 (Algebraic closure operator)**.** Let $\mathfrak{X}$ be a set. A map $f : 2^{\mathfrak{X}} \to 2^{\mathfrak{X}}$ is an *algebraic closure operator* if it satisfies extension, monotonicity, idempotence, and the finitary property: if $c \in f(\mathcal{K})$, then $c \in f(\mathcal{S})$ for some finite $\mathcal{S} \subseteq \mathcal{K}$.

Proof of Theorem 2.15. For (i), extension, monotonicity, and idempotence of $\mathrm{cl}_{\mathfrak{m}}$ follow from Proposition 2.10. Finitariness follows from Theorem 2.14 and Lemma A.2: if $c \in \mathrm{cl}_{\mathfrak{m}}(\mathcal{K})$, then there is a finite derivation tree using only finitely many base labels from $\mathcal{K}$.

For (ii), define $\mathcal{E} = \{(\mathcal{S}, c) : \mathcal{S} \subseteq \mathfrak{X} \text{ finite and } c \in f(\mathcal{S}) \setminus \mathcal{S}\}$. Let g be the closure operator induced by $\mathcal{E}$ as in the theorem statement. We show $g(\mathcal{K}) = f(\mathcal{K})$ for every $\mathcal{K} \subseteq \mathfrak{X}$.

First, $g(\mathcal{K}) \subseteq f(\mathcal{K})$ because $f(\mathcal{K})$ is a fixed point of $\Psi_{\mathcal{E}}$ containing $\mathcal{K}$: if $c \in \Psi_{\mathcal{E}}(f(\mathcal{K}))$, then either $c \in f(\mathcal{K})$ or else there exists $(\mathcal{S}, c) \in \mathcal{E}$ with $\mathcal{S} \subseteq f(\mathcal{K})$, which implies $c \in f(\mathcal{S}) \subseteq f(f(\mathcal{K})) = f(\mathcal{K})$.

Conversely, if $c \in f(\mathcal{K})$, then by algebraicity there exists a finite $\mathcal{S}_0 \subseteq \mathcal{K}$ such that $c \in f(\mathcal{S}_0)$. If $c \in \mathcal{S}_0$, then $c \in \mathcal{K} \subseteq g(\mathcal{K})$. Otherwise $(\mathcal{S}_0, c) \in \mathcal{E}$, and since $\mathcal{S}_0 \subseteq \mathcal{K} \subseteq g(\mathcal{K})$, the rule fires inside $g(\mathcal{K})$, so $c \in g(\mathcal{K})$. □

Proof of Theorem 2.19

If $c^* \in \mathcal{A}_{\mathfrak{m}}$, the empty curriculum works. Assume therefore that $c^* \notin \mathcal{A}_{\mathfrak{m}}$. Since $c^* \in \mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$, Theorem 2.14 implies that there exists a derivation tree of c^* from $\mathcal{A}_{\mathfrak{m}}$. By Lemma A.2, this derivation tree is finite. Let $\mathcal{R}$ be the set of all non-base rule nodes in this derivation tree. Form a directed graph on $\mathcal{R}$ by retaining the parent-child relation between rule nodes and orienting each edge from child to parent. Because the derivation tree is finite and well-founded, this directed graph is finite and acyclic. Hence it admits a topological ordering $v_1, \ldots, v_L$. For each $i = 1, \ldots, L$, let $(\mathcal{S}_i, c_i)$ be the expansion rule attached to the node v_i. Define $\mathcal{K}_0 = \mathcal{A}_{\mathfrak{m}}$, $\mathcal{K}_i = \mathcal{K}_{i-1} \cup \{c_i\}$ for $i = 1, \ldots, L$. We claim that $\gamma = ((\mathcal{S}_1, c_1), \ldots, (\mathcal{S}_L, c_L))$ is a valid ordered curriculum starting from $\mathcal{A}_{\mathfrak{m}}$. Indeed, fix $i \in \{1, \ldots, L\}$ and let $s \in \mathcal{S}_i$. In the derivation tree, the child corresponding to s is either:

(i) a base node, in which case $s \in \mathcal{A}_{\mathfrak{m}} = \mathcal{K}_0 \subseteq \mathcal{K}_{i-1}$; or

(ii) a rule node. In that case this child must occur earlier than v_i in the topological order, say it is v_j with $j < i$. Its label is then $c_j = s$, so $s \in \mathcal{K}_j \subseteq \mathcal{K}_{i-1}$. Thus every prerequisite in $\mathcal{S}_i$ belongs to $\mathcal{K}_{i-1}$, so $\mathcal{S}_i \subseteq \mathcal{K}_{i-1}$. Since also $(\mathcal{S}_i, c_i) \in \mathcal{E}_{\mathfrak{m}}$ by construction, each step is valid. Therefore γ is a valid ordered curriculum. Finally, the root of the derivation tree is a rule node labelled by c^*. Hence it is one of the nodes $v_1, \ldots, v_L$, say v_r, and therefore $c_r = c^*$. It follows that $c^* \in \mathcal{K}_r \subseteq \mathcal{K}_L$. So the curriculum reaches a final acquired concept set containing c^*.

Proof of Proposition 2.21

We argue by induction on i. For $i = 0$, $\mathcal{K}_0 = \mathcal{A}_{\mathfrak{m}} \subseteq \mathcal{U}_{\mathfrak{m}}$ by Proposition 2.11(i). Now suppose $\mathcal{K}_{i-1} \subseteq \mathcal{U}_{\mathfrak{m}}$. Since γ is valid, $(\mathcal{S}_i, c_i) \in \mathcal{E}_{\mathfrak{m}}$ and $\mathcal{S}_i \subseteq \mathcal{K}_{i-1} \subseteq \mathcal{U}_{\mathfrak{m}}$. By Proposition 2.11(ii), this implies $c_i \in \mathcal{U}_{\mathfrak{m}}$. Hence $\mathcal{K}_i = \mathcal{K}_{i-1} \cup \{c_i\} \subseteq \mathcal{U}_{\mathfrak{m}}$. This proves the claim for all i. The final statement follows immediately.

Proof of Proposition 2.24

Because $\mathcal{U}_{\mathfrak{m}}$ is finite by Assumption 2.23, the power set $2^{\mathcal{U}_{\mathfrak{m}}}$ is finite. Since $\mathbb{K}_{\mathfrak{m}} \subseteq 2^{\mathcal{U}_{\mathfrak{m}}}$, it follows that $\mathbb{K}_{\mathfrak{m}}$ is finite.

For (i), the trivial chain of length zero shows that $\mathcal{A}_{\mathfrak{m}} \in \mathbb{K}_{\mathfrak{m}}$. Moreover, every reachable set contains $\mathcal{A}_{\mathfrak{m}}$, since every witnessing chain starts from $\mathcal{A}_{\mathfrak{m}}$ and only adds concepts. Thus $\mathcal{A}_{\mathfrak{m}}$ is the minimum element of $(\mathbb{K}_{\mathfrak{m}}, \subseteq)$.

For (ii), let $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}}$ with $\mathcal{K} \neq \mathcal{A}_{\mathfrak{m}}$. By definition of reachability, there exists a witnessing chain

$$\mathcal{A}_{\mathfrak{m}} = \mathcal{K}_0 \subset \mathcal{K}_1 \subset \cdots \subset \mathcal{K}_L = \mathcal{K}$$

such that

$$\mathcal{K}_{i+1} = \mathcal{K}_i \cup \{c_i\}, \qquad c_i \in \Phi_{\mathfrak{m}}(\mathcal{K}_i) \setminus \mathcal{K}_i \qquad (i = 0, \ldots, L-1).$$

Since $\mathcal{K} \neq \mathcal{A}_{\mathfrak{m}}$, one has $L \geq 1$. Then $\mathcal{K}_{L-1} \in \mathbb{K}_{\mathfrak{m}}$ by Lemma B.2,

$$\mathcal{K}_{L-1} \subset \mathcal{K}, \qquad |\mathcal{K} \setminus \mathcal{K}_{L-1}| = 1.$$

This proves (ii).

For (iii), let $\mathcal{K}, \mathcal{K}' \in \mathbb{K}_{\mathfrak{m}}$. Choose a witnessing chain for $\mathcal{K}'$:

$$\mathcal{A}_{\mathfrak{m}} = \mathcal{K}'_0 \subset \mathcal{K}'_1 \subset \cdots \subset \mathcal{K}'_s = \mathcal{K}', \qquad \mathcal{K}'_{i+1} = \mathcal{K}'_i \cup \{c_i\}, \qquad c_i \in \Phi_{\mathfrak{m}}(\mathcal{K}'_i) \setminus \mathcal{K}'_i.$$

For each $i = 0, \ldots, s$, define

$$\mathcal{L}_i = \mathcal{K} \cup \mathcal{K}'_i.$$

Then $\mathcal{L}_0 = \mathcal{K}$ and $\mathcal{L}_s = \mathcal{K} \cup \mathcal{K}'$. Since $\mathcal{K}'_i \subseteq \mathcal{L}_i$, monotonicity of $\Phi_{\mathfrak{m}}$ gives

$$\Phi_{\mathfrak{m}}(\mathcal{K}'_i) \subseteq \Phi_{\mathfrak{m}}(\mathcal{L}_i).$$

Hence, whenever $c_i \notin \mathcal{L}_i$, one has $c_i \in \Phi_{\mathfrak{m}}(\mathcal{K}'_i) \subseteq \Phi_{\mathfrak{m}}(\mathcal{L}_i)$, so

$$\mathcal{L}_{i+1} = \mathcal{L}_i \cup \{c_i\}$$

is a valid extension. If instead $c_i \in \mathcal{L}_i$, then $\mathcal{L}_{i+1} = \mathcal{L}_i$.

Removing repeated sets from the sequence $(\mathcal{L}_i)_{i=0}^{s}$ yields a valid chain from $\mathcal{K}$ to $\mathcal{K} \cup \mathcal{K}'$. Concatenating this chain with any witnessing chain from $\mathcal{A}_{\mathfrak{m}}$ to $\mathcal{K}$ shows that $\mathcal{K} \cup \mathcal{K}' \in \mathbb{K}_{\mathfrak{m}}$. Thus $\mathbb{K}_{\mathfrak{m}}$ is union-closed.

For (iv), we first show that $\mathcal{U}_{\mathfrak{m}} \in \mathbb{K}_{\mathfrak{m}}$. Let $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}}$ with $\mathcal{K} \neq \mathcal{U}_{\mathfrak{m}}$. Suppose, toward a contradiction, that $\Phi_{\mathfrak{m}}(\mathcal{K}) = \mathcal{K}$. Then $\mathcal{K}$ is a fixed point of $\Phi_{\mathfrak{m}}$ containing $\mathcal{A}_{\mathfrak{m}}$. Since $\mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ is the least fixed point containing $\mathcal{A}_{\mathfrak{m}}$, it follows that $\mathcal{U}_{\mathfrak{m}} \subseteq \mathcal{K}$. But by definition of $\mathbb{K}_{\mathfrak{m}}$, one also has $\mathcal{K} \subseteq \mathcal{U}_{\mathfrak{m}}$, hence $\mathcal{K} = \mathcal{U}_{\mathfrak{m}}$, a

contradiction. Therefore $\Phi_{\mathfrak{m}}(\mathcal{K}) \setminus \mathcal{K} \neq \varnothing$. Choose any $c \in \Phi_{\mathfrak{m}}(\mathcal{K}) \setminus \mathcal{K}$. Because $\mathcal{K} \subseteq \mathcal{U}_{\mathfrak{m}}$ and $\mathcal{U}_{\mathfrak{m}}$ is a fixed point of $\Phi_{\mathfrak{m}}$, monotonicity gives $\Phi_{\mathfrak{m}}(\mathcal{K}) \subseteq \Phi_{\mathfrak{m}}(\mathcal{U}_{\mathfrak{m}}) = \mathcal{U}_{\mathfrak{m}}$, so in particular $c \in \mathcal{U}_{\mathfrak{m}}$. Hence $\mathcal{K} \cup \{c\}$ is again a reachable subset of $\mathcal{U}_{\mathfrak{m}}$.

Starting from $\mathcal{A}_{\mathfrak{m}}$, repeat this step as long as the current reachable set is not equal to $\mathcal{U}_{\mathfrak{m}}$. Because $\mathcal{U}_{\mathfrak{m}}$ is finite and each step strictly enlarges the set, the process terminates after finitely many steps at some reachable set $\mathcal{F} \subseteq \mathcal{U}_{\mathfrak{m}}$ satisfying $\Phi_{\mathfrak{m}}(\mathcal{F}) = \mathcal{F}$. Since $\mathcal{F}$ is a fixed point containing $\mathcal{A}_{\mathfrak{m}}$, minimality of $\mathcal{U}_{\mathfrak{m}} = \mathrm{cl}_{\mathfrak{m}}(\mathcal{A}_{\mathfrak{m}})$ implies $\mathcal{U}_{\mathfrak{m}} \subseteq \mathcal{F}$. As also $\mathcal{F} \subseteq \mathcal{U}_{\mathfrak{m}}$, we conclude that $\mathcal{F} = \mathcal{U}_{\mathfrak{m}}$. Therefore $\mathcal{U}_{\mathfrak{m}} \in \mathbb{K}_{\mathfrak{m}}$. Since every element of $\mathbb{K}_{\mathfrak{m}}$ is by definition a subset of $\mathcal{U}_{\mathfrak{m}}$, it follows that $\mathcal{U}_{\mathfrak{m}}$ is the maximum element of $(\mathbb{K}_{\mathfrak{m}}, \subseteq)$.

Finally, (v) follows from (iii). For any $\mathcal{K}, \mathcal{K}' \in \mathbb{K}_{\mathfrak{m}}$, the set $\mathcal{K} \cup \mathcal{K}'$ belongs to $\mathbb{K}_{\mathfrak{m}}$ and is clearly an upper bound of $\mathcal{K}$ and $\mathcal{K}'$. If $\mathcal{M} \in \mathbb{K}_{\mathfrak{m}}$ is any other upper bound, so that

$$\mathcal{K} \subseteq \mathcal{M} \qquad \text{and} \qquad \mathcal{K}' \subseteq \mathcal{M},$$

then $\mathcal{K} \cup \mathcal{K}' \subseteq \mathcal{M}$. Hence $\mathcal{K} \cup \mathcal{K}'$ is the least upper bound. □

Proof of Corollary 2.26

The result follows directly from Proposition 2.24 and Definition 2.25. □

Proof of Theorem 2.27

We prove (ii)⇒(i) and (i)⇒(ii) separately.

(ii)⇒(i). Assume there exists a mind $\mathfrak{m} = (\mathcal{C}, \mathcal{A}, \mathcal{E}_{\mathfrak{m}})$ such that $\mathbb{K}_{\mathfrak{m}} = \mathbb{F}$. By Corollary 2.26, the family $\mathbb{K}_{\mathfrak{m}}$ is an $\mathcal{A}$-based learning space. Hence so is $\mathbb{F}$.

(i)⇒(ii). Assume that $\mathbb{F}$ is an $\mathcal{A}$-based learning space, and define $\mathfrak{m}_{\mathbb{F}} = (\mathcal{C}, \mathcal{A}, \mathcal{E}_{\mathbb{F}})$ using the canonical rule set above. Let $\mathbb{K}_{\mathfrak{m}_{\mathbb{F}}}$ denote the reachable family generated by this mind. We prove that $\mathbb{K}_{\mathfrak{m}_{\mathbb{F}}} = \mathbb{F}$. *Step 1:* $\mathbb{K}_{\mathfrak{m}_{\mathbb{F}}} \subseteq \mathbb{F}$. Let $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}_{\mathbb{F}}}$. Choose a witnessing chain $\mathcal{A} = \mathcal{K}_0 \subset \mathcal{K}_1 \subset \cdots \subset \mathcal{K}_L = \mathcal{K}$ such that

$$\mathcal{K}_{i+1} = \mathcal{K}_i \cup \{c_i\}, \qquad c_i \in \Phi_{\mathfrak{m}_{\mathbb{F}}}(\mathcal{K}_i) \setminus \mathcal{K}_i \quad \text{for } i = 0, \ldots, L-1.$$

We prove by induction on i that $\mathcal{K}_i \in \mathbb{F}$ for all i.

For $i = 0$, one has $\mathcal{K}_0 = \mathcal{A} \in \mathbb{F}$.

Now suppose $\mathcal{K}_i \in \mathbb{F}$. Since $c_i \in \Phi_{\mathfrak{m}_{\mathbb{F}}}(\mathcal{K}_i) \setminus \mathcal{K}_i$, there exists a rule $(\mathcal{S}, c_i) \in \mathcal{E}_{\mathbb{F}}$ with $\mathcal{S} \subseteq \mathcal{K}_i$. By definition of $\mathcal{E}_{\mathbb{F}}$, $\mathcal{S} \in \mathbb{F}$, $\mathcal{S} \cup \{c_i\} \in \mathbb{F}$. Because $\mathcal{K}_i \in \mathbb{F}$ and $\mathbb{F}$ is union-closed,

$$\mathcal{K}_i \cup (\mathcal{S} \cup \{c_i\}) \in \mathbb{F}.$$

Since $\mathcal{S} \subseteq \mathcal{K}_i$, this simplifies to

$$\mathcal{K}_i \cup \{c_i\} = \mathcal{K}_{i+1} \in \mathbb{F}.$$

Thus every $\mathcal{K}_i$ lies in $\mathbb{F}$, and in particular $\mathcal{K} \in \mathbb{F}$. Hence $\mathbb{K}_{\mathfrak{m}_{\mathbb{F}}} \subseteq \mathbb{F}$. *Step 2:* $\mathbb{F} \subseteq \mathbb{K}_{\mathfrak{m}_{\mathbb{F}}}$. Let $\mathcal{K} \in \mathbb{F}$. If $\mathcal{K} = \mathcal{A}$, then $\mathcal{K}$ is reachable by the trivial chain.

Assume now that $\mathcal{K} \neq \mathcal{A}$. Since $\mathbb{F}$ is an $\mathcal{A}$-based learning space, repeated application of accessibility yields a descending chain

$$\mathcal{K} = \mathcal{K}_L \supset \mathcal{K}_{L-1} \supset \cdots \supset \mathcal{K}_0 = \mathcal{A}$$

such that each $\mathcal{K}_i \in \mathbb{F}$ and

$$\mathcal{K}_i = \mathcal{K}_{i-1} \cup \{x_i\} \qquad \text{for } i = 1, \ldots, L.$$

Reverse the chain:

$$\mathcal{A} = \mathcal{K}_0 \subset \mathcal{K}_1 \subset \cdots \subset \mathcal{K}_L = \mathcal{K}.$$

For each $i = 1, \ldots, L$, both $\mathcal{K}_{i-1}$ and $\mathcal{K}_i = \mathcal{K}_{i-1} \cup \{x_i\}$ belong to $\mathbb{F}$. Therefore, by the definition of $\mathcal{E}_{\mathbb{F}}$, $(\mathcal{K}_{i-1}, x_i) \in \mathcal{E}_{\mathbb{F}}$. Hence $x_i \in \Phi_{\mathfrak{m}_{\mathbb{F}}}(\mathcal{K}_{i-1}) \setminus \mathcal{K}_{i-1}$, so every step in the chain is a valid reachable extension. Thus $\mathcal{K}$ is reachable from $\mathcal{A}$, which shows that $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}_{\mathbb{F}}}$. Therefore $\mathbb{F} \subseteq \mathbb{K}_{\mathfrak{m}_{\mathbb{F}}}$.

Combining the two inclusions gives $\mathbb{K}_{\mathfrak{m}_{\mathbb{F}}} = \mathbb{F}$. □

A.2 Proofs for Section 3

Proof of Lemma 3.5

We argue by induction on t.

For $t = 0$, $\mathcal{K}_0 = \mathcal{A}_{\mathfrak{m}} \in \mathbb{K}_{\mathfrak{m}}$ by the trivial witnessing chain. Now suppose $\mathcal{K}_t \in \mathbb{K}_{\mathfrak{m}}$ almost surely. If $Y_{t+1} = \perp$, then by Definition 3.4, $\mathcal{K}_{t+1} = \mathcal{K}_t$, hence $\mathcal{K}_{t+1} \in \mathbb{K}_{\mathfrak{m}}$. If $Y_{t+1} \in \mathcal{Z}$, define $c_{t+1} = \mathrm{tgt}(Y_{t+1})$. Since the parser outputs a non-null signal, Definition 3.3 implies that $c_{t+1} \in \Phi_{\mathfrak{m}}(\mathcal{K}_t)$. Hence either $c_{t+1} \in \mathcal{K}_t$, in which case $\mathcal{K}_{t+1} = \mathcal{K}_t$, or else $c_{t+1} \in \Phi_{\mathfrak{m}}(\mathcal{K}_t) \setminus \mathcal{K}_t$, in which case $\mathcal{K}_{t+1} = \mathcal{K}_t \cup \{c_{t+1}\}$ is a valid one-step reachable extension from $\mathcal{K}_t$ in the sense of Definition 2.22. Since $\mathcal{K}_t \in \mathbb{K}_{\mathfrak{m}}$, it follows that $\mathcal{K}_{t+1} \in \mathbb{K}_{\mathfrak{m}}$. This proves the claim. □

Proof of Proposition 3.10

By the definition of conditional mutual information,

$$\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t) = \mathbb{H}(\Theta \mid \mathcal{F}_t) - \mathbb{E}[\mathbb{H}(\Theta \mid \mathcal{F}_t \vee \sigma(Y_{t+1})) \mid \mathcal{F}_t].$$

Since $\mathcal{F}_{t+1} = \mathcal{F}_t \vee \sigma(Y_{t+1})$ and $H_t = \mathbb{H}(\Theta \mid \mathcal{F}_t)$, this becomes $\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t) = H_t - \mathbb{E}[H_{t+1} \mid \mathcal{F}_t]$. Because H_t is $\mathcal{F}_t$-measurable, $H_t - \mathbb{E}[H_{t+1} \mid \mathcal{F}_t] = \mathbb{E}[H_t - H_{t+1} \mid \mathcal{F}_t]$. □

Proof of Theorem 3.11

By Proposition 3.10,

$$H_t - \mathbb{E}[H_{t+1} \mid \mathcal{F}_t] = \mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t) \geq 0.$$

Hence (H_t) is a supermartingale. Equality holds if and only if

$$\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t) = 0,$$

which is equivalent to conditional independence of Θ and Y_{t+1} given $\mathcal{F}_t$. □

Proof of Theorem 3.13

The proof of Theorem 3.13 relies on the following lemma, which we establish first.

Lemma A.4 (Unparseability erases information)**.** *Let*

$$C_{t+1} = \mathrm{tgt}(Z_{t+1}), \qquad U_{t+1} = \{C_{t+1} \notin \Phi_{\mathfrak{m}}(\mathcal{K}_t)\}.$$

Then:

(i) *on* U_{t+1} *one has* $Y_{t+1} = \perp$ *almost surely, and therefore*

$$\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t, U_{t+1}) = 0, \qquad \mathbb{I}(Z_{t+1}; Y_{t+1} \mid \mathcal{F}_t, U_{t+1}) = 0;$$

(ii) *if* $\mathbb{P}(U_{t+1} \mid \mathcal{F}_t) = 1$*, then*

$$\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t) = 0, \qquad \mathbb{I}(Z_{t+1}; Y_{t+1} \mid \mathcal{F}_t) = 0.$$

Proof. On U_{t+1}, Definition 3.3 gives

$$Y_{t+1} = \rho_{\mathfrak{m}}(Z_{t+1}, \mathcal{K}_t) = \perp \qquad \text{almost surely.}$$

Hence conditional on $(\mathcal{F}_t, U_{t+1})$, the random variable Y_{t+1} is constant, so all the relevant conditional entropies are zero. This proves (i).

If $\mathbb{P}(U_{t+1} \mid \mathcal{F}_t) = 1$, then U_{t+1} occurs almost surely conditional on $\mathcal{F}_t$, so $Y_{t+1} = \perp$ almost surely conditional on $\mathcal{F}_t$. Again all relevant conditional entropies are zero, proving (ii). □

Proof of Theorem 3.13. On U_{t+1}, Lemma A.4 gives $\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t, U_{t+1}) = 0$. On U_{t+1}^c, Proposition B.3 yields $\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t, U_{t+1}^c) = \mathbb{I}(\Theta; Z_{t+1} \mid \mathcal{F}_t, U_{t+1}^c) > 0$. □

A.3 Proofs for Section 4

Lemma A.5 (Explicit formula for the parsed entropy bound)**.** *Assume* $\mathcal{Z}$ *is finite. Then, for every* $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}}$*,*

$$C_{\mathfrak{m}}(\mathcal{K}) = \begin{cases} \log\Big(|\mathcal{Z}_{\mathrm{ord}}(\mathcal{K})| + 1\Big), & \text{if } \mathcal{Z}_{\mathrm{ord}}(\mathcal{K}) \subsetneq \mathcal{Z}, \\ \log|\mathcal{Z}|, & \text{if } \mathcal{Z}_{\mathrm{ord}}(\mathcal{K}) = \mathcal{Z}. \end{cases}$$

Proof of Lemma A.5. Fix $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}}$ and define the parsed observation range

$$\mathcal{Y}(\mathcal{K}) = \{\rho_{\mathfrak{m}}(z, \mathcal{K}) : z \in \mathcal{Z}\} \subseteq \mathcal{Z} \cup \{\perp\}.$$

For any $\mathcal{Z}$-valued random variable Z, the random variable $\rho_{\mathfrak{m}}(Z, \mathcal{K})$ takes values in $\mathcal{Y}(\mathcal{K})$ almost surely, so $\mathbb{H}(\rho_{\mathfrak{m}}(Z, \mathcal{K})) \le \log|\mathcal{Y}(\mathcal{K})|$ by [Cover and Thomas, 2006, Page 41]. Taking the supremum over all such Z gives $C_{\mathfrak{m}}(\mathcal{K}) \le \log|\mathcal{Y}(\mathcal{K})|$.

For the reverse inequality, let $M = |\mathcal{Y}(\mathcal{K})|$. For each $y \in \mathcal{Y}(\mathcal{K})$, choose some representative $z_y \in \mathcal{Z}$ such that $\rho_{\mathfrak{m}}(z_y, \mathcal{K}) = y$. Define a $\mathcal{Z}$-valued random variable Z by

$$\mathbb{P}(Z = z_y) = \frac{1}{M} \qquad \text{for each } y \in \mathcal{Y}(\mathcal{K}),$$

and $\mathbb{P}(Z = z) = 0$ for all other $z \in \mathcal{Z}$. Then $\rho_{\mathfrak{m}}(Z, \mathcal{K})$ is uniform on $\mathcal{Y}(\mathcal{K})$, so

$$\mathbb{H}(\rho_{\mathfrak{m}}(Z, \mathcal{K})) = \log |\mathcal{Y}(\mathcal{K})|.$$

Hence $C_{\mathfrak{m}}(\mathcal{K}) = \log |\mathcal{Y}(\mathcal{K})|$.

Under the parsing map $\rho_{\mathfrak{m}}$, one has

$$\rho_{\mathfrak{m}}(z, \mathcal{K}) = \begin{cases} z, & \text{if } z \in \mathcal{Z}_{\mathrm{ord}}(\mathcal{K}), \\ \bot, & \text{if } z \notin \mathcal{Z}_{\mathrm{ord}}(\mathcal{K}). \end{cases}$$

If $\mathcal{Z}_{\mathrm{ord}}(\mathcal{K}) \subsetneq \mathcal{Z}$, then $\mathcal{Y}(\mathcal{K}) = \mathcal{Z}_{\mathrm{ord}}(\mathcal{K}) \cup \{\bot\}$, so $|\mathcal{Y}(\mathcal{K})| = |\mathcal{Z}_{\mathrm{ord}}(\mathcal{K})| + 1$. If instead $\mathcal{Z}_{\mathrm{ord}}(\mathcal{K}) = \mathcal{Z}$, then every raw signal is parseable, so $\mathcal{Y}(\mathcal{K}) = \mathcal{Z}$. Substituting these two cases into $C_{\mathfrak{m}}(\mathcal{K}) = \log |\mathcal{Y}(\mathcal{K})|$ proves the claim. □

Proof of Proposition 4.3

Because $\mathcal{K}_t$ is $\mathcal{F}_t$-measurable, conditional on $\mathcal{F}_t$ the law of $Y_{t+1} = \rho_{\mathfrak{m}}(Z_{t+1}, \mathcal{K}_t)$ is obtained by passing the conditional law of Z_{t+1} through the fixed map $z \mapsto \rho_{\mathfrak{m}}(z, \mathcal{K}_t)$. Therefore,

$$\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t) \le \mathbb{H}(Y_{t+1} \mid \mathcal{F}_t) \le C_{\mathfrak{m}}(\mathcal{K}_t) \qquad \text{almost surely.}$$

□

Proof of Lemma 4.4

Assume $\mathcal{K}, \mathcal{K}' \in \mathbb{K}_{\mathfrak{m}}$ with $\mathcal{K} \subseteq \mathcal{K}'$. By Lemma 2.7, $\Phi_{\mathfrak{m}}(\mathcal{K}) \subseteq \Phi_{\mathfrak{m}}(\mathcal{K}')$, hence $\mathcal{Z}_{\mathrm{ord}}(\mathcal{K}) \subseteq \mathcal{Z}_{\mathrm{ord}}(\mathcal{K}')$.

Define $g_{\mathcal{K},\mathcal{K}'} : \mathcal{Z} \cup \{\bot\} \to \mathcal{Z} \cup \{\bot\}$ by

$$g_{\mathcal{K},\mathcal{K}'}(y) = \begin{cases} y, & \text{if } y \in \mathcal{Z}_{\mathrm{ord}}(\mathcal{K}), \\ \bot, & \text{otherwise.} \end{cases}$$

Then for every $z \in \mathcal{Z}$, $\rho_{\mathfrak{m}}(z, \mathcal{K}) = g_{\mathcal{K},\mathcal{K}'}(\rho_{\mathfrak{m}}(z, \mathcal{K}'))$. Indeed, if $z \in \mathcal{Z}_{\mathrm{ord}}(\mathcal{K})$, then z is ordered at both sets and both sides equal z. If $z \notin \mathcal{Z}_{\mathrm{ord}}(\mathcal{K})$, then the left-hand side is $\bot$; on the right-hand side, either $\rho_{\mathfrak{m}}(z, \mathcal{K}') = \bot$, or else $\rho_{\mathfrak{m}}(z, \mathcal{K}') = z$ and $g_{\mathcal{K},\mathcal{K}'}(z) = \bot$. Now let Z be any $\mathcal{Z}$-valued random variable. Then

$$\rho_{\mathfrak{m}}(Z, \mathcal{K}) = g_{\mathcal{K},\mathcal{K}'}\big(\rho_{\mathfrak{m}}(Z, \mathcal{K}')\big) \qquad \text{almost surely.}$$

Thus $\rho_{\mathfrak{m}}(Z, \mathcal{K})$ is a deterministic function of $\rho_{\mathfrak{m}}(Z, \mathcal{K}')$. By the data processing inequality, $\mathbb{H}(\rho_{\mathfrak{m}}(Z, \mathcal{K})) \le \mathbb{H}(\rho_{\mathfrak{m}}(Z, \mathcal{K}'))$. Taking suprema over all $\mathcal{Z}$-valued random variables Z yields $C_{\mathfrak{m}}(\mathcal{K}) \le C_{\mathfrak{m}}(\mathcal{K}')$. □

Proof of Theorem 4.6

Let $g_{\mathcal{K},\mathcal{K}'}$ be the deterministic map constructed in the proof of Lemma 4.4. For every raw signal $z \in \mathcal{Z}$, one has $\rho_{\mathfrak{m}}(z,\mathcal{K}) = g_{\mathcal{K},\mathcal{K}'}(\rho_{\mathfrak{m}}(z,\mathcal{K}'))$. Therefore, conditional on the public history h_t,

$$\rho_{\mathfrak{m}}(Z_{t+1},\mathcal{K}) = g_{\mathcal{K},\mathcal{K}'}\Big(\rho_{\mathfrak{m}}(Z_{t+1},\mathcal{K}')\Big) \qquad \text{almost surely.}$$

Hence, for every $\omega \in \Omega$ and every $y \in \mathcal{Z} \cup \{\perp\}$,

$$\mathbb{W}_{\mathcal{K},h_t}(y \mid \omega) = \sum_{y' \in \mathcal{Z}\cup\{\perp\}} G_{\mathcal{K},\mathcal{K}'}(y \mid y')\, \mathbb{W}_{\mathcal{K}',h_t}(y' \mid \omega),$$

where $G_{\mathcal{K},\mathcal{K}'}(y \mid y') = \mathbf{1}\{g_{\mathcal{K},\mathcal{K}'}(y') = y\}$ is the Markov kernel induced by $g_{\mathcal{K},\mathcal{K}'}$.

Thus $\mathbb{W}_{\mathcal{K},h_t}$ is obtained from $\mathbb{W}_{\mathcal{K}',h_t}$ by post-processing through a Markov kernel independent of ω. Therefore $\mathbb{W}_{\mathcal{K},h_t}$ is a garbling of $\mathbb{W}_{\mathcal{K}',h_t}$, and $\mathbb{W}_{\mathcal{K}',h_t}$ Blackwell-dominates $\mathbb{W}_{\mathcal{K},h_t}$. □

Proof of Proposition 4.8

Fix a sample path. By the concept-acquisition update rule, each round adds at most one new concept to the learner's acquired concept set. If completion occurs at time τ, then in particular $\Theta \in \mathcal{K}_\tau$. Delete repeated sets from the sequence $\mathcal{K}_0, \mathcal{K}_1, \ldots, \mathcal{K}_\tau$. The resulting strictly increasing sequence is of the form

$$\mathcal{A}_{\mathfrak{m}} = \mathcal{K}_{i_0} \subset \mathcal{K}_{i_1} \subset \cdots \subset \mathcal{K}_{i_r},$$

where each step adds one concept belonging to the one-step expansion of the previous set. Hence it is a prerequisite-respecting chain ending at a set containing Θ.

By definition of $L_{\mathfrak{m}}(\Theta)$, any such chain has length at least $L_{\mathfrak{m}}(\Theta)$. Since the number of strict acquisitions up to time τ is at most τ, it follows that $\tau \geq L_{\mathfrak{m}}(\Theta)$ almost surely. Taking expectations completes the proof. □

Proof of Lemma 4.9

By Proposition 3.10, $\mathbb{E}[H_t - H_{t+1} \mid \mathcal{F}_t] = \mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t)$. Multiplying by $\mathbf{1}_{\{\tau_{\mathrm{id}} > t\}}$ and taking expectations gives

$$\mathbb{E}\left[\mathbf{1}_{\{\tau_{\mathrm{id}}>t\}}(H_t - H_{t+1})\right] = \mathbb{E}\left[\mathbf{1}_{\{\tau_{\mathrm{id}}>t\}}\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t)\right].$$

Summing from $t = 0$ to $n-1$ yields

$$\mathbb{E}\left[\sum_{t=0}^{n-1} \mathbf{1}_{\{\tau_{\mathrm{id}}>t\}}\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t)\right] = \mathbb{E}[H_0] - \mathbb{E}[H_{\tau_{\mathrm{id}}\wedge n}].$$

Since Θ is $\mathcal{F}_{\tau_{\mathrm{id}}}$-measurable, one has

$$H_{\tau_{\mathrm{id}}} = \mathbb{H}(\Theta \mid \mathcal{F}_{\tau_{\mathrm{id}}}) = 0 \qquad \text{almost surely.}$$

Also, $0 \le H_t \le \log|\Omega|$ for all t. Let

$$S_n = \sum_{t=0}^{n-1} \mathbf{1}_{\{\tau_{\mathrm{id}} > t\}} \mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t).$$

Because the summands are nonnegative, S_n increases almost surely to

$$\sum_{t=0}^{\tau_{\mathrm{id}}-1} \mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t).$$

Monotone convergence and bounded convergence therefore give

$$\mathbb{E}\left[\sum_{t=0}^{\tau_{\mathrm{id}}-1} \mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t)\right] = \mathbb{E}[H_0].$$

Since $\mathcal{F}_0$ is trivial, $\mathbb{E}[H_0] = \mathbb{H}(\Theta)$. □

Proof of Proposition 4.10

By Lemma 4.9,

$$\mathbb{H}(\Theta) = \mathbb{E}\left[\sum_{t=0}^{\tau_{\mathrm{id}}-1} \mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t)\right].$$

By Proposition 4.3,

$$\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t) \le C_{\mathfrak{m}}(\mathcal{K}_t) \qquad \text{almost surely for every } t.$$

Substituting this bound inside the sum yields the result. □

Proof of Theorem 4.11

By Proposition 4.8, the structural bound follows: $\mathbb{E}[\tau] \ge \mathbb{E}[L_{\mathfrak{m}}(\Theta)]$. For the epistemic part: define the identification time

$$\tau_{\mathrm{id}} = \inf\{t \ge 0 : \mathbb{H}(\Theta \mid \mathcal{F}_t) = 0\}.$$

Since $\{\tau_{\mathrm{id}} \le t\} = \{H(\Theta \mid \mathcal{F}_t) = 0\} \in \mathcal{F}_t$, τ_{id} is an $(\mathcal{F}_t)$-stopping time.

Moreover, if τ is a completion time then identification must already have occurred, so $\tau_{\mathrm{id}} \le \tau$ almost surely. By Proposition 4.10,

$$\mathbb{H}(\Theta) \le \mathbb{E}\left[\sum_{t=0}^{\tau_{\mathrm{id}}-1} C_{\mathfrak{m}}(\mathcal{K}_t)\right].$$

Since $\mathcal{K}_t \in \mathbb{K}_{\mathfrak{m}}$ almost surely by Lemma 3.5,

$$C_{\mathfrak{m}}(\mathcal{K}_t) \le C_{\mathfrak{m}}^{\max} \qquad \text{almost surely.}$$

Therefore

$$\mathbb{H}(\Theta) \le \mathbb{E}\left[\sum_{t=0}^{\tau_{\mathrm{id}}-1} C_{\mathfrak{m}}(\mathcal{K}_t)\right] \le C_{\mathfrak{m}}^{\max}\, \mathbb{E}[\tau_{\mathrm{id}}] \le C_{\mathfrak{m}}^{\max}\, \mathbb{E}[\tau].$$

Rearranging yields

$$\mathbb{E}[\tau] \ge \frac{\mathbb{H}(\Theta)}{C_{\mathfrak{m}}^{\max}}.$$

Combining the two bounds gives the theorem. □

Proof of Proposition 4.13

For each $c \in \Omega_+$, choose one raw signal $z_c \in \mathcal{Z}$ such that $\mathrm{tgt}(z_c) = c$. Because tgt is a function, the signals $(z_c)_{c\in\Omega_+}$ are pairwise distinct. Since $\Omega \subseteq \mathcal{U}_{\mathfrak{m}}$ and $\mathcal{U}_{\mathfrak{m}}$ is a fixed point of $\Phi_{\mathfrak{m}}$, every target concept $c \in \Omega$ is ordered at $\mathcal{U}_{\mathfrak{m}}$. Hence each z_c is parseable at $\mathcal{U}_{\mathfrak{m}}$, so the parsed observation range at $\mathcal{U}_{\mathfrak{m}}$ contains at least the distinct symbols $\{z_c : c \in \Omega_+\}$. Therefore, by Lemma A.5, $C^{\max}_{\mathfrak{m}} \geq C_{\mathfrak{m}}(\mathcal{U}_{\mathfrak{m}}) \geq \log|\Omega_+|$. Since entropy is bounded by the logarithm of the support size, $\mathbb{H}(\Theta) \leq \log|\Omega_+|$, which yields $\mathbb{H}(\Theta) \leq C^{\max}_{\mathfrak{m}}$. This proves (i).

For (ii), fix $c \in \Omega_+$. By definition of $L_{\mathfrak{m}}(c)$, there exists a valid ordered curriculum of length $L_{\mathfrak{m}}(c)$ from $\mathcal{A}_{\mathfrak{m}}$ to a set containing c. Under Assumption 4.12, the teacher can implement that curriculum by sending one raw signal targeting each concept along the path. After $L_{\mathfrak{m}}(c)$ rounds, the learner has acquired c.

In one additional round, the teacher sends the fixed representative signal z_c. Because $c \in \mathcal{K}_{L_{\mathfrak{m}}(c)}$, the signal z_c is parseable at that state. Since the strategy specifies a unique representative signal for each possible target, the learner identifies the realized target after observing z_c. Thus

$$\tau \leq L_{\mathfrak{m}}(\Theta) + 1 \qquad \text{almost surely.}$$

Taking expectations completes the proof. □

A.4 Proofs for Section 5

Proof of Proposition 5.1

By Proposition 4.8, every completion time τ satisfies

$$\tau \geq L_{\mathfrak{m}}(\Theta) \qquad \text{almost surely.}$$

Hence $\{\tau \leq t\} \subseteq \{L_{\mathfrak{m}}(\Theta) \leq t\}$. Therefore, for every admissible teaching strategy,

$$\mathbb{P}(\tau \leq t) \leq \mathbb{P}\Big(L_{\mathfrak{m}}(\Theta) \leq t\Big).$$

Taking the supremum over strategies yields the first claim.

If $t < L_{\min}$, then $L_{\mathfrak{m}}(\Theta) > t$ almost surely under the prior, so

$$\mathbb{P}\Big(L_{\mathfrak{m}}(\Theta) \leq t\Big) = 0.$$

Hence $\mathbb{V}(t) = 0$. □

Proof of Proposition 5.2

Fix an admissible teaching strategy with $\mathbb{E}[\tau] < \infty$. By Markov's inequality,

$$\mathbb{P}(\tau > t) \leq \frac{\mathbb{E}[\tau]}{t},$$

and therefore

$$\mathbb{P}(\tau \leq t) \geq 1 - \frac{\mathbb{E}[\tau]}{t}.$$

Since $\mathbb{V}(t)$ is the supremum of $\mathbb{P}(\tau \leq t)$ over all admissible strategies, it follows that

$$\mathbb{V}(t) \geq 1 - \frac{\mathbb{E}[\tau]}{t}.$$

Letting $t \to \infty$ gives $\mathbb{V}(t) \to 1$. □

Proof of Proposition 5.3

By Proposition 4.8, every acquisition time τ_g satisfies

$$\tau_g \geq L_{\mathfrak{m}}(g) \qquad \text{almost surely.}$$

Therefore, for every admissible strategy and every $t < L_{\mathfrak{m}}(g)$, $\mathbb{P}(\tau_g \leq t) = 0$. Taking the supremum over strategies yields $\mathbb{V}_g(t) = 0$ for $t < L_{\mathfrak{m}}(g)$.

Now let $L = L_{\mathfrak{m}}(g)$. By definition of structural distance, there exists a witnessing chain

$$\mathcal{A}_{\mathfrak{m}} = \mathcal{K}_0 \subset \mathcal{K}_1 \subset \cdots \subset \mathcal{K}_L, \qquad g \in \mathcal{K}_L,$$

such that

$$\mathcal{K}_{i+1} = \mathcal{K}_i \cup \{u_i\}, \qquad u_i \in \Phi_{\mathfrak{m}}(\mathcal{K}_i) \setminus \mathcal{K}_i \quad (i = 0, \ldots, L-1).$$

By Assumption 4.12, for each u_i there exists a raw signal $z_i \in \mathcal{Z}$ such that

$$\mathrm{tgt}(z_i) = u_i.$$

Since $u_i \in \Phi_{\mathfrak{m}}(\mathcal{K}_i)$, the signal z_i is parseable at $\mathcal{K}_i$. If the teacher sends

$$z_0, z_1, \ldots, z_{L-1}$$

in sequence, the learner moves through the sets

$$\mathcal{K}_0, \mathcal{K}_1, \ldots, \mathcal{K}_L$$

and therefore acquires g after L rounds. Thus there exists an admissible strategy such that

$$\mathbb{P}(\tau_g \leq L) = 1.$$

Hence

$$\mathbb{V}_g(t) = 1 \qquad \text{for all } t \geq L_{\mathfrak{m}}(g).$$

□

Proof of Proposition 5.5

For (i), if every learner receives fewer than L rounds, then by Proposition 5.3 the acquisition probability of g is zero for every learner. Hence no learner completes.

For (ii), select $\lfloor B/L \rfloor$ learners and allocate L rounds to each of them, allocating 0 rounds to the remaining learners. This is feasible because $L \lfloor B/L \rfloor \leq B$. By Proposition 5.3, each selected learner completes with probability one. Hence the total number of completed learners is $\lfloor B/L \rfloor$. □

Proof of Theorem 5.6

Let a and g be distinct concepts. For each $i \in \{1, \dots, k\}$ and each $j \in \{1, \dots, L-1\}$, let $p_{i,j}$ be a distinct concept, with all these concepts also distinct from a and g. Define

$$\mathcal{C} = \{a, g\} \cup \{p_{i,j} : i = 1, \dots, k,\ j = 1, \dots, L-1\}, \qquad \mathcal{A} = \{a\}.$$

For each $i \in \{1, \dots, k\}$, define the rule set of mind $\mathfrak{m}_i$ by

$$\mathcal{E}_{\mathfrak{m}_i} = \Big\{(\{a\}, p_{i,1}), (\{p_{i,1}\}, p_{i,2}), \dots, (\{p_{i,L-2}\}, p_{i,L-1}), (\{p_{i,L-1}\}, g)\Big\}.$$

Thus mind $\mathfrak{m}_i$ has a private prerequisite chain

$$a \to p_{i,1} \to p_{i,2} \to \cdots \to p_{i,L-1} \to g,$$

and no concept $p_{i',j}$ with $i' \neq i$ is reachable in mind $\mathfrak{m}_i$.

Choose raw signals

$$z_{i,j} \in \mathcal{Z} \qquad (i = 1, \dots, k,\ j = 1, \dots, L-1), \qquad z_g \in \mathcal{Z},$$

with

$$\mathrm{tgt}(z_{i,j}) = p_{i,j}, \qquad \mathrm{tgt}(z_g) = g.$$

(i) Personalized acquisition in L rounds. Fix i. The sequence

$$z_{i,1}, z_{i,2}, \dots, z_{i,L-1}, z_g$$

is a valid ordered curriculum of length L for $\mathfrak{m}_i$: each signal becomes parseable when its predecessor on the private chain has been acquired, and the final signal acquires g.

(ii) Broadcast lower bound. Consider any common broadcast sequence $\Gamma = (z_1, \dots, z_T)$ that acquires g for every mind. Fix i. Before mind $\mathfrak{m}_i$ can acquire g, it must first acquire all $L-1$ private prerequisite concepts

$$p_{i,1}, \dots, p_{i,L-1}.$$

Moreover, if $i' \neq i$, then none of the concepts $p_{i,j}$ lies in $\mathcal{U}_{\mathfrak{m}_{i'}}$. Hence a broadcast signal targeting $p_{i,j}$ can help at most mind $\mathfrak{m}_i$; it produces no acquisition for any other mind.

It follows that at least $L-1$ rounds must be devoted to the private prerequisites of each mind i. Summing over $i = 1, \dots, k$, at least $k(L-1)$ rounds are required to make all minds ready for a signal targeting g.

Finally, one additional round targeting g is necessary, since $g \notin \mathcal{A}$ and is acquired only when a signal with target g is parseable. Hence

$$T \geq k(L-1) + 1.$$

(iii) Tightness. Consider the broadcast sequence

$$z_{1,1}, \dots, z_{1,L-1}, z_{2,1}, \dots, z_{2,L-1}, \dots, z_{k,1}, \dots, z_{k,L-1}, z_g.$$

During the block $z_{i,1}, \dots, z_{i,L-1}$, only mind $\mathfrak{m}_i$ advances; all other minds ignore those signals. After the first $k(L-1)$ rounds, each mind $\mathfrak{m}_i$ has acquired $p_{i,L-1}$. The final signal z_g is therefore parseable for every mind, so all of them acquire g on the last round. Thus the lower bound is attained. □

B Additional results

This appendix collects supplementary results that are invoked in the proofs above but are not essential to the main narrative. It also records additional consequences of the framework that may be of independent interest.

Corollary B.1 (Not every union-closed family above the axioms is a learning space)**.** *The class of $\mathcal{A}$-based learning spaces on $\mathcal{U}$ is a strict subclass of the class of union-closed families $\mathbb{F} \subseteq 2^{\mathcal{U}}$.*

Proof. Every $\mathcal{A}$-based learning space is, by definition, union-closed and lies above $\mathcal{A}$, so only strictness needs to be shown. Let

$$\mathcal{U} = \{a, b\}, \qquad \mathcal{A} = \varnothing, \qquad \mathbb{F} = \{\varnothing, \{a, b\}\}.$$

Then $\mathbb{F}$ is union-closed and contains $\mathcal{A}$. However, it fails accessibility, since neither

$$\{a, b\} \setminus \{a\} = \{b\} \qquad \text{nor} \qquad \{a, b\} \setminus \{b\} = \{a\}$$

belongs to $\mathbb{F}$. Hence $\mathbb{F}$ is not an $\mathcal{A}$-based learning space. □

Lemma B.2 (Prefix closure of reachable acquired concept sets)**.** *If $\mathcal{K} \in \mathbb{K}_{\mathfrak{m}}$ and $\mathcal{A}_{\mathfrak{m}} = \mathcal{K}_0 \subset \mathcal{K}_1 \subset \cdots \subset \mathcal{K}_L = \mathcal{K}$ is a witnessing chain, then every intermediate set $\mathcal{K}_i$ also belongs to $\mathbb{K}_{\mathfrak{m}}$.*

Proof. Each $\mathcal{K}_i$ is reachable from $\mathcal{A}_{\mathfrak{m}}$ by truncating the witnessing chain at step i. □

Proposition B.3 (Parseability preserves information)**.** *Let*

$$U_{t+1} = \{\mathrm{tgt}(Z_{t+1}) \notin \Phi_{\mathfrak{m}}(\mathcal{K}_t)\}.$$

Then

$$\mathbb{I}(\Theta; Y_{t+1} \mid \mathcal{F}_t, U_{t+1}^c) = \mathbb{I}(\Theta; Z_{t+1} \mid \mathcal{F}_t, U_{t+1}^c).$$

In particular, if the right-hand side is strictly positive, then so is the left-hand side.

Proof. On U_{t+1}^c, the parser acts as the identity, so

$$Y_{t+1} = Z_{t+1} \qquad \text{almost surely.}$$

The identity of the conditional mutual informations follows immediately. □

Corollary B.4 (Unlimited rephrasing can be useless under sharp parsing)**.** *Fix time t and a mind $\mathfrak{m}$. Let $U_t(c) = \{c \notin \Phi_{\mathfrak{m}}(\mathcal{K}_t)\}$. Let $(Z_{t+1}^{(j)})_{j \geq 1}$ be any family of $\mathcal{Z}$-valued random variables such that*

$$\mathrm{tgt}(Z_{t+1}^{(j)}) = c \qquad \textit{almost surely for every } j \geq 1,$$

and define $Y_{t+1}^{(j)} = \rho_{\mathfrak{m}}(Z_{t+1}^{(j)}, \mathcal{K}_t)$. Then for every $j \geq 1$, $\mathbb{I}(\Theta; Y_{t+1}^{(j)} \mid \mathcal{F}_t) = 0$ almost surely on $U_t(c)$.

Proof. On $U_t(c)$, the targeted concept is not ordered, so $Y_{t+1}^{(j)} = \perp$ almost surely. Hence $Y_{t+1}^{(j)}$ is conditionally constant given $\mathcal{F}_t$, so the conditional mutual information is zero. □

References

C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, 2006.

Gary S. Becker. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press, 1964.

Yoram Ben-Porath. The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4):352–365, 1967.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *International Conference on Machine Learning*, pages 41–48, 2009.

Claude Berge. *Hypergraphs: Combinatorics of finite sets*, volume 45. Elsevier, 1984.

David Blackwell. Comparison of experiments. In *Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102. University of California Press, 1951.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2006.

Flavio Cunha and James J. Heckman. The technology of skill formation. *American Economic Review*, 97(2):31–47, 2007.

Reinhard Diestel. *Graph Theory*, volume 173. Springer, 2024.

Jean-Paul Doignon and Jean-Calude Falmagne. *Knowledge Spaces*. Springer, 1999.

Jean-Paul Doignon and Jean-Claude Falmagne. Knowledge spaces and learning spaces. *arXiv:1511.06757*, 2015.

Jean-Paul Doignon and Jean-Claude Falmagne. Knowledge spaces and learning spaces. In *New Handbook of Mathematical Psychology, Volume 1: Foundations and Methodology*, pages 274–321. Cambridge University Press, 2016.

Sally A. Goldman and Michael J. Kearns. On the complexity of teaching. *Journal of Computer and System Sciences*, 50(1):20–31, 1995.

Bernhard Korte and László Lovász. Structural properties of greedoids. *Combinatorica*, 3(3):359–374, 1983.

Bernhard Korte, László Lovász, and Rainer Schrader. *Greedoids*, volume 4 of *Algorithms and Combinatorics*. Springer-Verlag, 1991.

Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

Claude E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 1949.

Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.

Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285–309, 1955.

Hassler Whitney. On the abstract properties of linear dependence. *American Journal of Mathematics*, 57(3):509–533, 1935.

Xiaojin Zhu, Ji Liu, and Manuel Lopes. No learner left behind: On the complexity of teaching multiple learners simultaneously. In *International Joint Conference on Artificial Intelligence*, pages 3588–3594, 2017.

Xiaojin Zhu, Adish Singla, Sandra Zilles, and Anna N Rafferty. An overview of machine teaching. *arXiv:1801.05927*, 2018.