SPANNING, NOT COUNTING: NOVELTY IN LANGUAGE GENERATION IN THE LIMIT, FROM FINITENESS TO COMPACTNESS

Anonymous authors
Paper under double-blind review

ABSTRACT

A generator can produce endlessly many distinct sentences and keep saying the same thing, and language generation in the limit, in the model of Kleinberg and Mullainathan, cannot tell the difference: it asks only that outputs be unseen. We ask that each output lie outside the span of everything said so far, for a span given by topics, a matroid, a consequence relation, or any monotone map that contains its argument. The right notion of a small set is then a finite core, a set spanned by finitely many of its own points: for the identity a finite set, for a finitary closure operator a set whose closure is compact in the lattice of closed sets. With finite cores in place of finite sets, the characterizations of Kleinberg and Mullainathan (NeurIPS 2024) and of Raman, Li and Tewari (COLT 2025) hold for every span, at every level when only the examples count as said, and uniformly when the generator's own outputs count too. The uniform price is free or fatal: as many examples as before, or no number at all, as for the paraphrases of one statement. At the other levels the generator's bets are the only obstruction, through a second mechanism that counting cannot see: poisoning, a bet valid for one language that together with examples of another spans it. Poisoning is impossible when the span is locally Noetherian, can be fatal beyond, as in a logic with explosion, and cannot be detected from inside single languages. Matroids, and several matroids at once, are the main instances.

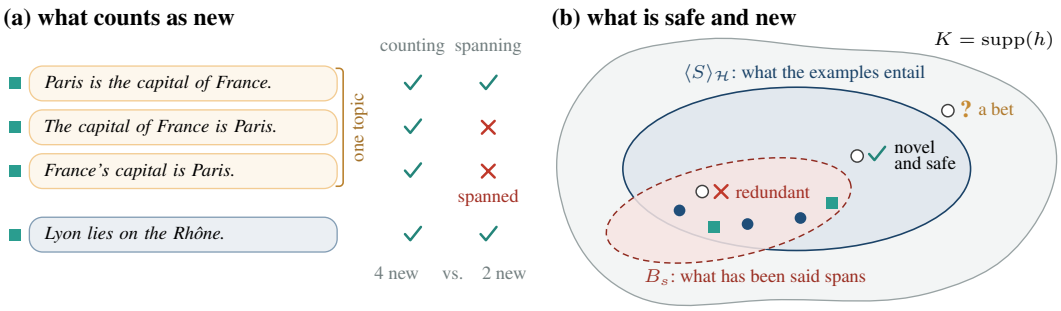

Figure 1: Counting versus spanning. (a) Four answers that never repeat a sentence all count as new when novelty counts strings; measured against the span of what has been said, here the partition of statements into topics, only two do. (b) Dots are examples, squares the generator's own outputs. The examples entail their safe zone $\langle S \rangle_{\mathcal{H}}$, and everything said spans $B_s = \Psi(E_s)$. An output that is novel and safe lies in the safe zone but outside the span; one inside the span is redundant, and one outside the safe zone is a bet.

1 INTRODUCTION

A generator that answers "Paris is the capital of France," then "The capital of France is Paris," then "France's capital is Paris" never repeats a sentence and never says anything new, and the theory of language generation in the limit cannot see the difference. In the model of Kleinberg & Mullainathan (2024), an adversary fixes an unknown language K from a known countable class and enumerates its strings one at a time; after each round a generator outputs a string, and it succeeds if from some

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

round on every output lies in K and has not yet been enumerated. Gold's classical theorem says that languages cannot, in general, be *identified* in the limit (Gold, 1967; Angluin, 1980); Kleinberg & Mullainathan (2024) showed that one can nevertheless always *generate* from them, for every countable class. Raman et al. (2025) characterized the uniform version by a *closure dimension*: it is possible exactly when $\mathrm{Cdim} < \infty$, and then $\mathrm{Cdim}+1$ distinct examples suffice. The blind spot is the success criterion: it counts. Recent work measures what else a generator does, through the density of its outputs (Kleinberg & Wei, 2025a;b), the trade-off between breadth and hallucination (Kalavasis et al., 2025; 2024; Charikar & Pabbaraju, 2025), the representation of subpopulations (Peale et al., 2025), or distance from the examples (Li et al., 2026a). None asks what the outputs span to keep growing.

Novelty relative to what has been said. Fix a *span*: a monotone map Ψ that sends a set E of strings to the strings $\Psi(E) \supseteq E$ that E makes redundant, and call an output *novel* if it lies outside the span of everything that has appeared so far, the adversary's examples *and* the generator's own earlier outputs (Figure 1). The identity recovers the original requirement, plus never repeating oneself. A partition into *topics*, say the paraphrases of each statement, makes a statement redundant once it has been made; of the three answers above, at most the first is novel. NoveltyBench counts outputs modulo a learned partition into functional-equivalence classes (Zhang et al., 2025), the role topics play here. A *linear* matroid asks for a new direction (in $\mathbb{Q}^d$ at most d times), as determinantal point processes and the Vendi score do softly (Kulesza & Taskar, 2012; Friedman & Dieng, 2023), and a *consequence relation* for a statement that does not follow from what has been said (Tarski, 1930). On the other side, the examples *entail* a string if every consistent language contains it: the entailed strings form the safe zone of Kleinberg & Mullainathan (2024). To be safe, a generator must say what the examples entail but what has been said does not span.

Finiteness was compactness. The classical theory counts: the closure dimension is the size of the largest set of examples whose safe zone is finite, and Kleinberg & Mullainathan (2024) generate from every countable class of infinite languages. For the identity span, a set is finite exactly when it is spanned by finitely many of its own points, when it has a *finite core*; for a finitary closure operator, a set has a finite core exactly when its closure is a compact element of the lattice of closed sets (Proposition 5.4; Davey & Priestley, 2002), and the compact elements of the lattice of all sets are the finite ones. Read every finite set of strings in the classical theory, a finite safe zone or a finite language, in this way, and the theory transfers verbatim: uniform generation is characterized by the *core dimension*, the size of the largest set of examples whose safe zone has a finite core, with one more example needed; non-uniform generation by chains of classes of finite core dimension; and, for countable classes, generation in the limit by languages without a finite core. The transfer holds

- at every level and for *every* span, when only the examples count as said (Proposition 5.1);
- when the generator's own outputs count too, uniformly for every span (Theorem 4.1), and at every level whenever the span is *locally Noetherian*: no set in the span of finitely many strings contains an infinite sequence of strings each outside the span of the ones before (Theorem 5.2).

The proofs at these levels are the classical ones, read through cores; what is new is the notion, and a second mechanism, poisoning, that counting cannot see. The identity recovers Kleinberg & Mullainathan (2024) and Raman et al. (2025) exactly, a matroid gives their analogues with rank in place of cardinality, and a union of matroid spans gives matroid intersection (Table 1 and Section 7).

Free or fatal. The translation never changes a finite answer: for every span,

$$\mathrm{Coredim}_{\Psi}(\mathcal{H}) \in \{\mathrm{Cdim}(\mathcal{H}), \infty\} \qquad \text{(Theorem 4.1)}.$$

Whenever uniform generation remains possible it needs exactly as many examples as under plain novelty; it becomes impossible exactly when some safe zone is infinite but spanned by finitely many of its own points. The paraphrases of one statement form such a set, spanned by any one of them. Take two languages that share these paraphrases and otherwise cover disjoint infinite sets of topics. A generator that paraphrases the shared statement is plainly novel from the first example; yet once a paraphrase has been shown, every novel output must leave the shared topic and so bets on one of the languages, while the adversary keeps showing paraphrases: no generator succeeds after any fixed number of examples, even one chosen per language (Example 4.4).

Table 1: A dictionary. A classical notion (Kleinberg & Mullainathan, 2024; Raman et al., 2025) (the optimality of $\mathrm{Cdim}+1$ is Corollary 4.2(i)) becomes, for a span Ψ, the notion in the middle, and for a matroid the one on the right. The uniform rows hold for every span (Theorem 4.1); the others hold for every span when only the examples count as said (Proposition 5.1), and for locally Noetherian spans when the generator's own outputs count too (Theorem 5.2).

	classical (identity)	any span Ψ	matroid
small set	finite	has a finite core	finite rank
dimension	Cdim	$\mathrm{Coredim}_\Psi \in \{\mathrm{Cdim}, \infty\}$	$\mathrm{RCdim}_M \in \{\mathrm{Cdim}, \infty\}$
uniform iff	$\mathrm{Cdim} < \infty$	$\mathrm{Coredim}_\Psi < \infty$	$\mathrm{RCdim}_M < \infty$
examples needed	$\mathrm{Cdim}+1$	$\mathrm{Coredim}_\Psi+1$	RCdim_M+1
non-uniform iff	finite-Cdim chain	finite-$\mathrm{Coredim}_\Psi$ chain	finite-RCdim_M chain
limit iff (countable)	always	no language with a finite core	all languages of infinite rank
extra hypothesis	none	locally Noetherian (own outputs)	none

Inside and outside. The examples always lie in the target, and a finite set inside a set without a finite core never spans it; this is why the examples alone never obstruct the transfer. The generator's own outputs are the only obstruction (Figure 5, Appendix F). Uniformly it can keep them inside what it knows, answering only from the safe zone. At the other levels it may have to bet before it knows, and a bet outside the target can then be *poisoned*: together with examples of the target, it spans the target from outside. A span is locally Noetherian exactly when spanning from anywhere and spanning from inside agree; this suffices for bets to do no harm (Theorem 5.2), and for a finitary closure space the rank form of the theory holds, uniformly and in the limit, for every class exactly then (Theorem 5.3). Beyond such spans, bets can be fatal: in a logic with explosion, a bet p followed by the example $\neg p$ is a contradiction, everything follows from what has been said, and no output is ever novel again (Example 5.5). They need not be: facing two master keys that together span everything, a generator can simply never say one, and nothing inside the languages tells the two cases apart (Propositions 5.6 and F.2).

Matroids, and several respects at once. Every matroid is locally Noetherian and the bases of its sets of finite rank are finite cores, so every characterization holds with rank in place of cardinality. The theories coincide at every level (in the limit, for countable classes) exactly when no infinite set has finite rank; otherwise one language separates them, and the two languages above separate non-uniform generation from the limit, which no countable class does under plain novelty (Theorems 6.1 and 6.2). Novelty in several respects at once, a new topic *and* a new style, is a union of matroid spans; a set has a finite core exactly when it holds no infinite placement of non-attacking rooks on the grid of topics and styles (Section 7).

Related work. Metric novelty (Li et al., 2026a) forbids a ball around each example, an examples-only span, though with covering numbers as its clock and possibly uncountable spaces. Covering a safe zone by balls centred anywhere does not decide uniform generation, even on a countable space; our cores are covers centred inside the set (Appendix H). Protocol variants (Bai et al., 2026; Mehrotra et al., 2025; Kleinberg et al., 2026a) change resources or targets, not, to our knowledge, the notion of novelty. Matroids and closed-set systems appear in learning from positive data, where the goal is identification (Gao et al., 2012; de Brecht et al., 2007) and global chain conditions such as finite elasticity appear (Wright, 1989; Motoki et al., 1991); for generation with its own outputs a local one suffices (Theorem 5.2), and only for its own bets. Apart from metric novelty, none of the work on generation that we cite measures novelty against a span.

2 Model

2.1 Generation in the limit, and entailment

Throughout, $\mathcal{X}$ is a countably infinite universe with a fixed enumeration $\mathcal{X} = \{\xi_1, \xi_2, \ldots\}$. A *hypothesis class* is a set $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$, and the *language* of $h \in \mathcal{H}$ is $\mathrm{supp}(h) = \{x : h(x) = 1\}$. As in Kleinberg & Mullainathan (2024); Raman et al. (2025), every language is infinite. For $Y \subseteq \mathcal{X}$ let $\mathcal{H}(Y) = \{h \in \mathcal{H} : Y \subseteq \mathrm{supp}(h)\}$ be the *version space* of Y. A finite S is *consistent* if $\mathcal{H}(S) \neq \emptyset$, and then it *entails* the strings of its *safe zone* $\langle S \rangle_{\mathcal{H}} := \bigcap_{h \in \mathcal{H}(S)} \mathrm{supp}(h)$ (Kleinberg &

Mullainathan, 2024), those that every language consistent with S contains. Entailment is a closure operator: $S \subseteq \langle S \rangle_{\mathcal{H}}$, the safe zone grows with S, and $S \subseteq T \subseteq \langle S \rangle_{\mathcal{H}}$ implies $\langle T \rangle_{\mathcal{H}} = \langle S \rangle_{\mathcal{H}}$ (Lemma 3.1(a)). We set $\langle S \rangle_{\mathcal{H}} = \bot$ for inconsistent S and write $\langle x_{1:n} \rangle_{\mathcal{H}}$ for the safe zone of $\{x_1, \ldots, x_n\}$.

A *generator* is a deterministic map $G : \mathcal{X}^* \to \mathcal{X}$; it receives no feedback. The adversary commits in advance to a hypothesis $h \in \mathcal{H}$ and a stream $x_1, x_2, \ldots$ of strings of $\mathrm{supp}(h)$: in the *limit game* a complete enumeration of $\mathrm{supp}(h)$, repetitions allowed; in the *uniform* and *non-uniform games* any such sequence. At round s the generator outputs $\hat{x}_s = G(x_{1:s})$. Its *history* is $E_s = \{x_1, \ldots, x_s\} \cup \{\hat{x}_1, \ldots, \hat{x}_{s-1}\}$, and $c_s = |\{x_1, \ldots, x_s\}|$ counts the distinct examples so far.

2.2 Spans

A *span* is a monotone map Ψ on subsets of $\mathcal{X}$ with $E \subseteq \Psi(E)$: the strings that E makes redundant.

Definition 2.1 (Novelty)**.** The *history span* at round s is $B_s := \Psi(E_s)$. The generator *succeeds* at round s if $\hat{x}_s \in \mathrm{supp}(h) \setminus B_s$: the output is valid and *novel*.

Plain novelty, the requirement of Kleinberg & Mullainathan (2024); Raman et al. (2025), asks only that $\hat{x}_s \in \mathrm{supp}(h) \setminus \{x_1, \ldots, x_s\}$; novel success implies it, since $B_s \supseteq E_s$.

Example 2.2 (Spans)**.** (i) The *identity* $\Psi(E) = E$: success is plain novelty plus never repeating oneself. (ii) A *finitary closure space* $(\mathcal{X}, \mathrm{cl})$: a span with $\mathrm{cl}(\mathrm{cl}(A)) = \mathrm{cl}(A)$ and $\mathrm{cl}(A) = \bigcup\{\mathrm{cl}(F) : F \subseteq A \text{ finite}\}$. The consequence operator of classical propositional logic, $\mathrm{Cn}(\Gamma) = \{\varphi : \Gamma \models \varphi\}$, is one (Tarski, 1930), finitary by the compactness theorem (Enderton, 2001). (iii) A *finitary matroid* M (Oxley, 2011): a finitary closure space with *exchange*, $y \in \mathrm{cl}_M(A \cup \{x\}) \setminus \mathrm{cl}_M(A) \Rightarrow x \in \mathrm{cl}_M(A \cup \{y\})$, and *rank* $r_M(A) := \min\{|F| : F \text{ finite}, \ A \subseteq \mathrm{cl}_M(F)\}$. In a *topic* matroid $\mathcal{X}$ is partitioned into topics and $\mathrm{cl}(A)$ is the union of the topics that A meets; in a *linear* matroid $\mathcal{X}$ lies in a rational vector space and $\mathrm{cl}(A) = \mathrm{span}_{\mathbb{Q}}(A) \cap \mathcal{X}$; the *free* matroid is the identity. (iv) Several spans at once, $\Psi(E) = \bigcup_{i \le k} \mathrm{cl}_{M_i}(E)$ (Section 7), or the ε-neighbourhood of a linear span.

Definition 2.3 (The three games)**.** Thresholds range over $d^\star \in \mathbb{N}_{\ge 1}$.

- $\mathcal{H}$ is *generatable in the limit* if some G achieves the following: for every $h \in \mathcal{H}$ and every complete enumeration of $\mathrm{supp}(h)$ there is t^* such that G succeeds at every round $s \ge t^*$.
- $\mathcal{H}$ is *uniformly generatable* if there are G and $d^\star$ such that for every $h \in \mathcal{H}$, every stream in $\mathrm{supp}(h)$, and every round t with $c_t \ge d^\star$, G succeeds at round t.
- $\mathcal{H}$ is *non-uniformly generatable* if there is G such that every $h \in \mathcal{H}$ has a threshold $d^\star(h)$ with the same guarantee.

All three depend on the span; for a matroid we also say *rank-novel*. The uniform guarantee binds from the *first* round at which the count reaches $d^\star$, even if the count then stays put, as in the proofs of Raman et al. (2025).

2.3 Inside and outside

The paper turns on where a spanning set lies. Call $A \subseteq \mathcal{X}$ *finitely dominated* if $A \subseteq \Psi(F)$ for some finite $F \subseteq \mathcal{X}$, and call a finite $C \subseteq A$ with $A \subseteq \Psi(C)$ a *finite core* of A: a dominating set may span A from outside, a core spans it from inside. For the identity both mean that A is finite; for a matroid both mean that A has finite rank (Lemma 3.1(d)). The span is *locally Noetherian*, or has *finite internal domination* (FID), if every finitely dominated set has a finite core; equivalently, no finitely dominated set contains an infinite sequence $o_1, o_2, \ldots$ with $o_i \notin \Psi(\{o_1, \ldots, o_{i-1}\})$ for all i (Proposition 5.4).

Definition 2.4 (Core dimension)**.** $\mathrm{Coredim}_\Psi(\mathcal{H})$ is the supremum of the sizes of the consistent sets of distinct strings whose safe zone has a finite core (0 if there is none). The *domination dimension* $\mathrm{Domdim}_\Psi(\mathcal{H})$ is defined in the same way with "finitely dominated" in place of "has a finite core".

For the identity both are the closure dimension $\mathrm{Cdim}(\mathcal{H})$ of Raman et al. (2025), the largest size of a consistent set with a finite safe zone; for a matroid both are its *rank-closure dimension* $\mathrm{RCdim}_M(\mathcal{H})$, the closure dimension with rank in place of cardinality. For a locally Noetherian span the two dimensions coincide.

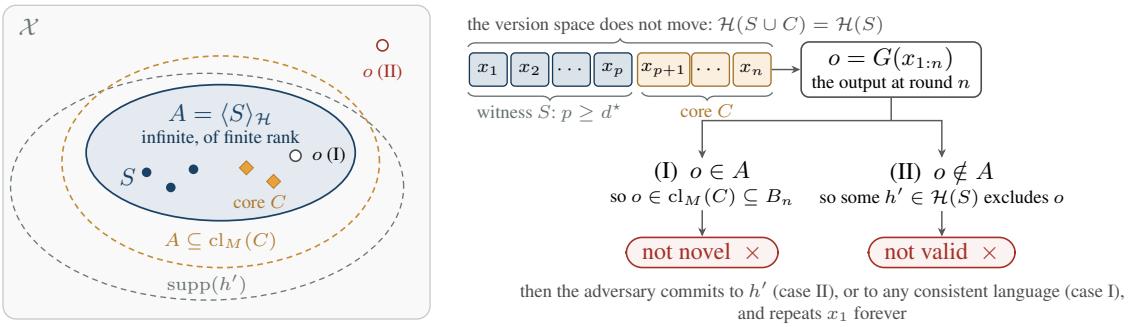

Figure 2: Exhaustion (Lemma 3.1(c)), drawn for a matroid; regions are schematic. The adversary shows a witness S as $x_1, \ldots, x_p$ and then a finite core C of its safe zone $A = \langle S \rangle_{\mathcal{H}}$ as $x_{p+1}, \ldots, x_n$; by part (a) the version space does not move. Whatever the generator outputs at round n is inside $A \subseteq \mathrm{cl}_M(C) \subseteq B_n$, and so not novel, or outside A, and so excluded by a consistent language.

3 One principle: escape or exhaustion

Every characterization applies one dichotomy to a well-chosen set A of candidate outputs, the safe zone or a language. In the next lemma, (a) is why the adversary can afford to show points of A, (b) and (c) are the two sides of the dichotomy, and (d) says what a finite core is for the three spans we use.

Lemma 3.1 (Escape or exhaustion)**.** *Let Ψ be a span, or any monotone map on subsets of $\mathcal{X}$, and let $B_s = \Psi(E_s)$.*

(a) (Showing the safe zone reveals nothing.) *If S and T are finite, $\mathcal{H}(S) \neq \emptyset$, and $S \subseteq T \subseteq \langle S \rangle_{\mathcal{H}}$, then $\mathcal{H}(T) = \mathcal{H}(S)$ and $\langle T \rangle_{\mathcal{H}} = \langle S \rangle_{\mathcal{H}}$.*

(b) (Escape.) *If A has no finite core, then $A \not\subseteq \Psi(E)$ for every finite $E \subseteq A$* (from inside)*. If A is not finitely dominated, then $A \not\subseteq \Psi(E)$ for every finite E* (from anywhere)*.*

(c) (Exhaustion.) *Let S be finite with $\mathcal{H}(S) \neq \emptyset$, and suppose $A = \langle S \rangle_{\mathcal{H}}$ has a finite core C. Then no deterministic generator achieves a uniform threshold $d^\star \leq |S|$, nor non-uniform thresholds with $d^\star(h) \leq |S|$ for every $h \in \mathcal{H}(S)$. Likewise, if a language $K = \mathrm{supp}(h)$ has a finite core, then on every complete enumeration of K no round succeeds once the core has appeared. Both statements hold even for the smaller span $\Psi(\{x_1, \ldots, x_s\})$ of the examples alone, which for Ψ the identity is plain novelty.*

(d) (Cores.) *For the identity, a set has a finite core iff it is finite; for a finitary matroid, iff it has finite rank, and then every basis of the set is a finite core; for $\Psi = \bigcup_{i \leq k} \mathrm{cl}_{M_i}$, iff it contains no infinite set that is independent in every M_i. Each of these three spans is locally Noetherian.*

Proof idea. (a) Every point of $\langle S \rangle_{\mathcal{H}}$ lies in every language consistent with S, so adding such points to S rules out no hypothesis. (b) restates the definitions, since the history is finite. For (c), the adversary presents S and then C (Figure 2). By (a) the version space never moves, and the threshold has been reached. The output o at the last round of this prefix is fixed, because the generator is deterministic. If $o \in A$, then $o \in \Psi(C)$ while C has been shown, so o is not novel; if $o \notin A$, some consistent hypothesis excludes o, and the adversary commits to it and repeats an example forever. (d) is where exchange enters: a maximal independent subset of a set of finite rank spans it. Proofs in Appendix B. □

Escape from inside is all a generator needs when it can keep its history inside A; escape from anywhere is what it needs when it cannot. For a locally Noetherian span the two agree. At the identity a set without a finite core is an infinite set, and the dichotomy is the classical one: infinite sets can be escaped, and finite ones are exhausted by showing them in full, as in the lower bound of Raman et al. (2025). A span moves the line from cardinality to cores. The new sets are the infinite ones with a finite core, such as an infinite topic or a plane, and exhaustion shows only the core.

4 UNIFORM GENERATION FOR EVERY SPAN

Throughout this section Ψ is a span.

Theorem 4.1 (Uniform generation for every span)**.** *A class $\mathcal{H}$ of any cardinality is uniformly generatable if and only if* $\mathrm{Coredim}_\Psi(\mathcal{H}) < \infty$*, and then exactly* $\mathrm{Coredim}_\Psi(\mathcal{H})+1$ *distinct examples are needed: this threshold is achievable, and no smaller one is. In particular, a generator that is valid and novel at every round exists if and only if* $\mathrm{Coredim}_\Psi(\mathcal{H}) = 0$*. Moreover* $\mathrm{Coredim}_\Psi(\mathcal{H}) \in \{\mathrm{Cdim}(\mathcal{H}), \infty\}$.

Proof idea. Exhaustion (Lemma 3.1(c)) is the lower bound: a witness with at least $d^\star$ strings whose safe zone has a finite core defeats threshold $d^\star$. The generator outputs a point of the current safe zone outside the history span, or any point of the zone if there is none. The history then stays inside the zone, and once the examples outnumber the core dimension the zone has no finite core, so the history never spans it (Lemma 3.1(b), from inside). Free or fatal: a finite safe zone is its own core, so $\mathrm{Cdim} \le \mathrm{Coredim}_\Psi$; an infinite zone with a finite core absorbs any number of further examples from itself without changing (Lemma 3.1(a)), so it gives witnesses of every size. Proofs in Appendix C. □

Beyond monotonicity, and extensivity for free or fatal, no structure of the span is used: not exchange, not idempotence, not finitarity. The theorem is the classical one for the identity and has the classical proof.

Corollary 4.2 (Three instances)**.** (i) *For the identity, uniform generation, plain or with repeats also forbidden, is possible if and only if* $\mathrm{Cdim}(\mathcal{H}) < \infty$ *(Raman et al., 2025), with optimal threshold* $\mathrm{Cdim}(\mathcal{H}) + 1$*; in particular a generator that is valid and unseen at every round exists if and only if* $\mathrm{Cdim}(\mathcal{H}) = 0$*.* (ii) *For a finitary matroid M the same holds with the rank-closure dimension* RCdim_M*, and* $\mathrm{RCdim}_M(\mathcal{H}) \in \{\mathrm{Cdim}(\mathcal{H}), \infty\}$*.* (iii) *For* $\Psi = \bigcup_{i \le k} \mathrm{cl}_{M_i}$ *the same holds with the common-independence dimension of Section 7.*

Rank novelty thus never changes what uniform generation costs, only whether it is possible.

Proposition 4.3 (One language separates the theories)**.** (i) *For every span, every $\mathcal{H}$ and each game, novel generatability implies plain generatability, with the same generator and thresholds.* (ii) *The converse fails already for one language: for the linear matroid on $\mathcal{X} = \mathbb{Q}^2$ and $\mathcal{H} = \{\mathbf{1}_{\mathbb{Q}^2}\}$, we have* $\mathrm{Cdim}(\mathcal{H}) = 0$*, so $\mathcal{H}$ is plainly generatable from the first example, while* $\mathrm{RCdim}_M(\mathcal{H}) = \infty$ *and $\mathcal{H}$ is not rank-novel generatable even in the limit.*

Two languages make the failure selective: fatal at every fixed threshold, harmless in the limit.

Example 4.4 (The paraphrase trap)**.** Partition $\mathcal{X}$ into infinite topics $T_0, T_1, T_2, \ldots$, the paraphrases of one statement each, with the topic matroid, and let $L_1 = T_0 \cup \bigcup_{i \text{ odd}} T_i$ and $L_2 = T_0 \cup \bigcup_{i \ge 2 \text{ even}} T_i$: two languages that share exactly one statement. For $\mathcal{H} = \{\mathbf{1}_{L_1}, \mathbf{1}_{L_2}\}$ every safe zone is T_0, L_1 or L_2, so $\mathrm{Cdim}(\mathcal{H}) = 0$, and paraphrasing T_0 is valid and unseen at every round. Under rank novelty, once an example from T_0 has been seen it spans T_0, so every later novel output lies in only one language, and exhaustion with the core $\{x_1\}$ defeats every pair of thresholds: $\mathcal{H}$ is not non-uniformly generatable. Both languages have infinite rank, so $\mathcal{H}$ is generatable in the limit (Theorem 5.2, Figure 3). The same holds for any two languages of infinite rank, in any finitary matroid, whose intersection is an infinite set of finite rank (Appendix E).

5 THE OTHER LEVELS: WHAT HAS BEEN SAID, AND LOCALLY NOETHERIAN SPANS

The examples always lie in the target; the generator's own outputs need not, and they count as said whether or not they are valid, since the generator cannot tell which are. Uniformly it can answer from the safe zone; at the other levels it may have to bet. A bet made for one language is *poisoned* when, together with examples of another, it spans that other language (Figure 5, Appendix F). This second mechanism is the one that counting cannot see.

Proposition 5.1 (Counting only the examples)**.** *Let Ψ be a span and measure novelty against the examples alone, $B_s = \Psi(\{x_1, \ldots, x_s\})$. Then $\mathcal{H}$ is uniformly generatable iff* $\mathrm{Coredim}_\Psi(\mathcal{H}) < \infty$*, with optimal threshold* $\mathrm{Coredim}_\Psi(\mathcal{H}) + 1$*; non-uniformly iff $\mathcal{H} = \bigcup_n \mathcal{H}_n$ for a chain $\mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq$*

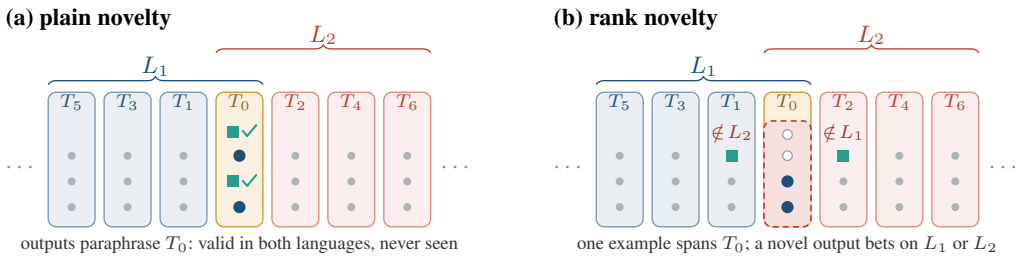

Figure 3: The paraphrase trap (Example 4.4): topics are columns, and L_1, L_2 share only T_0. (a) Paraphrasing T_0 is plainly novel forever. (b) Under rank novelty one example spans T_0, so a novel output bets on one language and the adversary commits to the other; only a complete enumeration, which eventually leaves T_0, reveals the target.

$\cdots$ with $\mathrm{Coredim}_\Psi(\mathcal{H}_n) < \infty$ for all n; and, if countable, in the limit iff no language has a finite core. With the generator's own outputs counted, $B_s = \Psi(E_s)$, the uniform answer is the same for every span, and the other two are the same (the limit for countable classes) whenever the span is locally Noetherian.

For the identity these are the characterizations of Kleinberg & Mullainathan (2024) and Raman et al. (2025); the proof is escape from inside at every level, since the examples lie in every safe zone and in every language the scaffold of Kleinberg & Mullainathan (2024) selects. Own outputs are the only obstruction; for the full history the transfer reads as follows.

Theorem 5.2 (Transfer with own outputs)**.** *Let Ψ be a span and $B_s = \Psi(E_s)$.*

(a) (Limit.) *Let $\mathcal{H}$ be countable. If no language of $\mathcal{H}$ is finitely dominated, $\mathcal{H}$ is generatable in the limit; if some language has a finite core, it is not. So for a locally Noetherian span, $\mathcal{H}$ is generatable in the limit iff no language has a finite core.*

(b) (Non-uniform.) *If $\mathcal{H}$ is non-uniformly generatable, then $\mathcal{H} = \bigcup_n \mathcal{H}_n$ for a chain with $\mathrm{Coredim}_\Psi(\mathcal{H}_n) < \infty$ for all n; conversely, such a chain with $\mathrm{Domdim}_\Psi(\mathcal{H}_n) < \infty$ for all n suffices. So for a locally Noetherian span, $\mathcal{H}$ is non-uniformly generatable iff such a chain exists, and a countable $\mathcal{H}$ is iff each of its finite subclasses has finite core dimension.*

For matroids, a countable class is generatable in the limit iff every language has infinite rank.

Proof idea. (a) The critical-language scaffold of Kleinberg & Mullainathan (2024), which uses no property of the span, eventually selects at each round a language of the class contained in the target; escape from anywhere yields a point of it outside the history span, whatever the generator said before. A language with a finite core is exhausted once its core has appeared. (b) A generator's thresholds sort the class into a chain of uniformly generatable subclasses, of finite core dimension by Theorem 4.1. Conversely, at round s the generator plays the uniform strategy for the largest $n \le s$ with $\mathrm{Domdim}_\Psi(\mathcal{H}_n) < c_s$; its earlier outputs may lie outside the current safe zone, which is why this direction needs escape from anywhere. Proofs in Appendix D. □

For a finitary closure space, the rank form of the theory, stated with finitely dominated sets, holds uniformly and in the limit for every class exactly when the space is locally Noetherian. We write $\mathrm{Domdim}_{\mathrm{cl}}$ for the domination dimension of cl.

Theorem 5.3 (The rank form needs exactly local Noetherianity)**.** *For a finitary closure space $(\mathcal{X}, \mathrm{cl})$ the following are equivalent:* (i) *it is locally Noetherian;* (ii) *for every class $\mathcal{H}$, $\mathcal{H}$ is uniformly generatable iff $\mathrm{Domdim}_{\mathrm{cl}}(\mathcal{H}) < \infty$, and then the optimal threshold is $\mathrm{Domdim}_{\mathrm{cl}}(\mathcal{H}) + 1$;* (iii) *for every countable class $\mathcal{H}$, $\mathcal{H}$ is generatable in the limit iff no language of $\mathcal{H}$ is finitely dominated. The sufficiency halves of (ii) and (iii) hold in every finitary closure space, and each necessity half, even restricted to one-language classes, already implies (i).*

Proof idea. Under local Noetherianity the two dimensions coincide and Theorems 4.1 and 5.2 apply. Conversely, if A is finitely dominated but has no finite core, it is infinite, and the one-language class $\{\mathbf{1}_A\}$ has $\mathrm{Domdim}_{\mathrm{cl}} = \infty$ and a finitely dominated language. Yet outputting the least point of A

outside the history span succeeds at every round: the history stays a finite subset of A, which never spans A (Lemma 3.1(b)). The finite sets that span A are not contained in it, so the adversary can never show them. Proofs in Appendix F. □

Proposition 5.4 (Locally Noetherian spans). (i) *A span is locally Noetherian iff no finitely dominated set contains an infinite sequence $o_1, o_2, \ldots$ with $o_i \notin \Psi(\{o_1, \ldots, o_{i-1}\})$ for all i.* (ii) *For a finitary closure operator, a set has a finite core iff its closure is finitely generated, and the span is locally Noetherian iff every increasing chain of closed sets inside any finitely generated closed set is eventually constant: the ascending chain condition (Noether, 1921) below every finitely generated closed set. Exchange and bounded such sequences are strictly stronger conditions (Proposition F.1).*

Example 5.5 (Explosion). Let the strings be the literals $p_j, \neg p_j$ ($j \geq 1$), and let a set of literals span every literal it entails in classical logic: itself if it contains no pair $\{p_j, \neg p_j\}$, and everything otherwise. The languages are the literals true under an assignment that makes only finitely many variables false. No language has a finite core, yet every language lies in the span of one pair: the span is not locally Noetherian. Every finite subclass is generatable from round one (Theorem 4.1), but the whole class is not generatable in the limit. An adversary that simulates the generator on the language of true literals finds its first success, with a literal p_j, and commits in advance to the language with $\neg p_j$, enumerated identically up to that round and then $\neg p_j$; from then on the history is inconsistent, spans everything, and no output is ever novel again. Counting only the examples as said, the class is generatable in the limit (Proposition 5.1).

Leaving local Noetherianity need not be fatal. With two master keys a, b that together span everything, while a set without both spans only itself, a generator that never says a key keeps the core form of Theorem 5.2 at every level (Proposition F.2). From inside the languages the two situations look the same.

Proposition 5.6 (Same inside, opposite answers). *Let $\mathcal{X}_0 = \{a, b, z_1, z_2, \ldots\}$ with $\mathrm{cl}_0(E) = \mathcal{X}_0$ if $\{a, b\} \subseteq E$ and $\mathrm{cl}_0(E) = E$ otherwise, and send the literals of Example 5.5 into it by $\varphi(p_j) = z_{2j-1}$, $\varphi(\neg p_j) = z_{2j}$. Then φ carries the languages of Example 5.5 onto those of a class $\mathcal{H}^\varphi$, and every set A inside a single language is its own span, as is $\varphi(A)$, and both are finitely dominated. With the generator's own outputs counted, $\mathcal{H}^\varphi$ is generatable in the limit and non-uniformly, and the class of Example 5.5 at neither level. So beyond locally Noetherian spans no criterion that sees only the languages and the sets inside single languages, up to such a relabelling, decides these levels.*

Proof idea. Neither space lets a set inside a language span more than itself, so φ preserves everything inside the languages; the answers are Example 5.5 and Proposition F.2. Proofs in Appendix F. □

6 Matroids: the hierarchy, and when the theories coincide

For a finitary matroid, Theorems 4.1 and 5.2 hold with rank in place of cardinality (Corollary 4.2 and Theorem 5.2). Two consequences have no classical counterpart: a finite class that separates non-uniform generation from the limit, and an exact criterion for when the two theories coincide.

Theorem 6.1 (Strict hierarchy). *Within rank-novel generation, uniform $\subseteq$ non-uniform $\subseteq$ limit, and both inclusions are strict for suitable matroids. The first gap is witnessed at the free matroid; the second by the* two-language *class of Example 4.4, whose languages intersect in a single infinite topic, and by two infinite-rank flats of a linear matroid that intersect in a plane. At the free matroid no finite class witnesses the second gap, since finite classes are uniformly generatable there.*

Proof idea. The first gap is classical. The second is Example 4.4: the adversary shows many paraphrases of the shared statement, and exhaustion defeats every pair of per-hypothesis thresholds; a complete enumeration, by contrast, eventually leaves the shared topic and reveals the target, and escape takes over. Two flats that intersect in a plane behave in the same way, with a basis of the plane as the core. Proofs in Appendix E. □

Theorem 6.2 (When the theories coincide). *For a finitary matroid M on $\mathcal{X}$ the following are equivalent:* (a) $\mathrm{RCdim}_M(\mathcal{H}) = \mathrm{Cdim}(\mathcal{H})$ *for every class $\mathcal{H}$ on $\mathcal{X}$;* (b) *every finite-rank flat of M is finite;* (c) *M has no infinite set of finite rank. Under these conditions every level of rank-novel*

generation coincides, as a property of classes, with the classical one (at the limit level, for countable classes), and every M violating them carries a one-language class realizing the separation of Proposition 4.3(ii).

Proof idea. If sets of finite rank are finite, cores are finite sets and the two dimensions have the same witnesses. Otherwise an infinite set of finite rank, as a one-language class, is plainly generatable from the first example and rank-novel generatable at no level. Proofs in Appendix E. □

7 Several respects at once: matroid intersection

Diversity is often required in several respects at once, say a new topic and a new style. Model each respect by a finitary matroid M_i ($i \leq k$, k finite) and take the span $\Psi(E) = \bigcup_{i \leq k} \mathrm{cl}_{M_i}(E)$: an output is novel if it is new in every respect. A set is *common independent* if it is independent in every M_i; the *common-independence capacity* $\rho_\cap(A)$ is the supremum of the sizes of its finite common independent subsets, and the *common-independence dimension* $\mathrm{RCdim}_\cap(\mathcal{H})$ is defined as RCdim_M with $\rho_\cap$ for r_M. For a topic and a style matroid, place each string on the grid of (topic, style) pairs: common independent sets are placements of non-attacking rooks. This span need not be idempotent, and matroid intersection is not a matroid, yet nothing more is needed.

Proposition 7.1 (Capacity). *If $A \subseteq \Psi(E)$ with E finite, every finite common independent $I \subseteq A$ has $|I| \leq k|E|$; if $\rho_\cap(A) < \infty$, a maximum common independent $I \subseteq A$ satisfies $A \subseteq \Psi(I)$. So A has a finite core iff it is finitely dominated iff it contains no infinite common independent set; the span is locally Noetherian, and its core dimension is $\mathrm{RCdim}_\cap \in \{\mathrm{Cdim}, \infty\}$, with $\mathrm{RCdim}_\cap \geq \max_i \mathrm{RCdim}_{M_i}$. By Theorems 4.1 and 5.2, $\mathcal{H}$ is uniformly generatable iff $\mathrm{RCdim}_\cap(\mathcal{H}) < \infty$, with threshold $\mathrm{RCdim}_\cap(\mathcal{H}) + 1$; non-uniformly iff $\mathcal{H}$ is the union of a chain of classes of finite $\mathrm{RCdim}_\cap$; and, if countable, in the limit iff every language contains an infinite common independent set.*

Proposition 7.2 (Rank in each matroid is not enough). *There are finitary matroids M_1, M_2 on a countable $\mathcal{X}$ and a one-language class $\mathcal{H} = \{h\}$ with $r_{M_1}(\mathrm{supp}\, h) = r_{M_2}(\mathrm{supp}\, h) = \infty$ and $\mathrm{supp}(h) = I_1 \cup I_2$, where I_i is independent in M_i, such that nevertheless $\rho_\cap(\mathrm{supp}\, h) = 2$ and $\mathcal{H}$ is not simultaneously generatable at any level.*

Proof idea. One inequality carries the theory: a common independent set meets each span $\mathrm{cl}_{M_i}(E)$ in at most $|E|$ points (Steinitz), hence $\Psi(E)$ in at most $k|E|$. So a set of infinite capacity is not finitely dominated, and a set of finite capacity is spanned by any maximum common independent subset (exchange in each M_i). Proposition 7.2 is the cross on the rook grid, one topic together with one style: once a string from each arm has been seen, every string repeats a topic or a style (Figure 7). Figure and proofs in Appendix G. □

8 Discussion

In short, to be safe a generator must say what the examples entail but what has been said does not span, and the classical theory of this task was a theory of compactness. It transfers to every span uniformly, and at every level when only the examples count as said; with the generator's own outputs it transfers whenever the span is locally Noetherian, and beyond such spans poisoning decides, invisibly from inside single languages (Proposition 5.6). For practice, novelty up to classes of equivalent outputs, as NoveltyBench measures it, costs no extra examples whenever uniform generation remains possible, and no fixed number of examples suffices as soon as two plausible targets share an infinite class. Like its predecessors, the model is worst-case and deterministic, and it adds a known span. Under a matroid of meanings, the three answers about Paris count once.

Open problems. (1) *Beyond locally Noetherian spans.* Bets in a fixed region that no examples can poison are harmless (Proposition F.3), but no such region need exist (Example F.4). Is generation in the limit with own outputs exactly the ability to confine bets to regions that grow with the examples and eventually cannot be poisoned (Appendix F)? (2) *Infinitely many constraints.* Proposition 7.1 uses $k < \infty$ only in the bound $|I| \leq k|E|$; what governs countably many matroids? (3) *Uncountable classes.* Plain novelty is characterized there by finite witnesses (Li et al., 2026b); do cores play their role for spans?

AI USE STATEMENT

The original research ideas and initial manuscript draft were developed by the authors. AI tools were used to assist with checking mathematical details and the correctness of arguments, filling in routine proof steps, drawing TikZ figures, and language polishing. The authors take full responsibility for the content and correctness of the final manuscript.

REPRODUCIBILITY STATEMENT

All assumptions are stated in Section 2, and complete proofs of every result are in the appendix, which opens with a map from each result to its proof. The finite checks of the explicit constructions are described there and provided as a script in the supplementary material.

REFERENCES

Antonios Anastasopoulos, Giuseppe Ateniese, and Evgenios M. Kornaropoulos. Safe language generation in the limit. *arXiv preprint arXiv:2601.08648*, 2026.

Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45(2):117–135, 1980.

Marcelo Arenas, Pablo Barceló, Luis Cofré, and Alexander Kozachinskiy. Language generation: Complexity barriers and implications for learning. *arXiv preprint arXiv:2511.05759*, 2025.

Marta Arias, José L. Balcázar, and Cristina Tîrnăucă. Learning definite horn formulas from closure queries. *arXiv preprint arXiv:1503.09025*, 2015.

Yannan Bai, Debmalya Panigrahi, and Ian Zhang. Language generation in the limit: Noise, loss, and feedback. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2026. arXiv:2507.15319.

Moses Charikar and Chirag Pabbaraju. Exploring facets of language generation in the limit. In *Proceedings of the 38th Annual Conference on Learning Theory (COLT)*, volume 291 of *Proceedings of Machine Learning Research*, 2025. arXiv:2411.15364.

Collette R. Coullard and Lisa Hellerstein. Independence and port oracles for matroids, with an application to computational learning theory. *Combinatorica*, 16(2):189–208, 1996. doi: 10.1007/BF01844845.

B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, second edition, 2002. ISBN 978-0-521-78451-1. doi: 10.1017/CBO9780511809088.

Matthew de Brecht, Masanori Kobayashi, Hiroo Tokunaga, and Akihiro Yamamoto. Inferability of closed set systems from positive data. In *New Frontiers in Artificial Intelligence (JSAI 2006 Workshops)*, volume 4384 of *Lecture Notes in Computer Science*. Springer, 2007. doi: 10.1007/978-3-540-69902-6_23.

Herbert B. Enderton. *A Mathematical Introduction to Logic*. Harcourt/Academic Press, second edition, 2001. ISBN 978-0-12-238452-3.

Nicolas Flammarion, Chirag Pabbaraju, Hristo Papazov, Miltiadis Stouras, and Ola Svensson. Space-efficient language generation in the limit. *arXiv preprint arXiv:2606.25777*, 2026.

Dan Friedman and Adji Bousso Dieng. The vendi score: A diversity evaluation metric for machine learning. *Transactions on Machine Learning Research*, 2023. arXiv:2210.02410.

Pierre Gabriel. Des catégories abéliennes. *Bulletin de la Société Mathématique de France*, 90: 323–448, 1962. doi: 10.24033/bsmf.1583.

Atul Ganju, Travis McVoy, Shaddin Dughmi, and Shang-Hua Teng. A theory of time-sensitive language generation: Sparse hallucination beats mode collapse. *arXiv preprint arXiv:2605.11302*, 2026.

Ziyuan Gao, Frank Stephan, Guohua Wu, and Akihiro Yamamoto. Learning families of closed sets in matroids. In *Computation, Physics and Beyond (WTCS 2012)*, volume 7160 of *Lecture Notes in Computer Science*, pp. 120–139. Springer, 2012. doi: 10.1007/978-3-642-27654-5_10.

E. Mark Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

Steve Hanneke, Amin Karbasi, Anay Mehrotra, and Grigoris Velegkas. On union-closedness of language generation. *arXiv preprint arXiv:2506.18642*, 2025.

Alkis Kalavasis, Anay Mehrotra, and Grigoris Velegkas. On characterizations for language generation: Interplay of hallucinations, breadth, and stability. *arXiv preprint arXiv:2412.18530*, 2024.

Alkis Kalavasis, Anay Mehrotra, and Grigoris Velegkas. On the limits of language generation: Trade-offs between hallucination and mode collapse. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC)*, 2025. doi: 10.1145/3717823.3718108. arXiv:2411.09642.

Amin Karbasi, Omar Montasser, John Sous, and Grigoris Velegkas. (im)possibility of automated hallucination detection in large language models. *arXiv preprint arXiv:2504.17004*, 2025.

Jon Kleinberg and Sendhil Mullainathan. Language generation in the limit. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. arXiv:2404.06757.

Jon Kleinberg and Fan Wei. Density measures for language generation. *arXiv preprint arXiv:2504.14370*, 2025a.

Jon Kleinberg and Fan Wei. Language generation and identification from partial enumeration: Tight density bounds and topological characterizations. *arXiv preprint arXiv:2511.05295*, 2025b.

Jon Kleinberg and Fan Wei. Banach density of generated languages: Dichotomies in topology and dimension. *arXiv preprint arXiv:2604.02385*, 2026.

Jon Kleinberg, Anay Mehrotra, Amin Saberi, and Grigoris Velegkas. On language generation in the limit with bounded memory. *arXiv preprint arXiv:2605.30324*, 2026a.

Jon Kleinberg, Charlotte Peale, and Omer Reingold. Mistake-bounded language generation. *arXiv preprint arXiv:2605.10809*, 2026b.

Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning*, 5(2–3):123–286, 2012. doi: 10.1561/2200000044.

Aaron Li and Ian Zhang. Characterizing the effect of noise in language generation in the limit. *arXiv preprint arXiv:2601.21237*, 2026.

Jiaxun Li, Vinod Raman, and Ambuj Tewari. On generation in metric spaces. *arXiv preprint arXiv:2602.07710*, 2026a.

Xiaoyu Li, Andi Han, Jiaojiao Jiang, and Junbin Gao. Characterizing language generation in the limit: Finite witnesses and a separation-width hierarchy. *arXiv preprint arXiv:2609.10525*, 2026b.

Xiaoyu Li, Andi Han, Jiaojiao Jiang, and Junbin Gao. Contrastive identification and generation in the limit. *arXiv preprint arXiv:2605.06211*, 2026c.

Anay Mehrotra, Grigoris Velegkas, Xifan Yu, and Felix Zhou. Language generation with infinite contamination. *arXiv preprint arXiv:2511.07417*, 2025.

Anay Mehrotra, Grigoris Velegkas, Xifan Yu, and Felix Zhou. Differentially private language generation and identification in the limit. *arXiv preprint arXiv:2604.08504*, 2026.

Tatsuya Motoki, Takeshi Shinohara, and Keith Wright. The correct definition of finite elasticity: Corrigendum to identification of unions. In *Proceedings of the Fourth Annual Workshop on Computational Learning Theory (COLT)*, pp. 375, 1991. doi: 10.1016/B978-1-55860-213-7.50037-7.

Zafeiria Moumoulidou, Andrew McGregor, and Alexandra Meliou. Diverse data selection under fairness constraints. *arXiv preprint arXiv:2010.09141*, 2020.

Emmy Noether. Idealtheorie in Ringbereichen. *Mathematische Annalen*, 83:24–66, 1921. doi: 10.1007/BF01464225.

James Oxley. *Matroid Theory*. Oxford University Press, second edition, 2011.

Charlotte Peale, Vinod Raman, and Omer Reingold. Representative language generation. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, 2025. arXiv:2505.21819.

Giorgio Racca, Michal Valko, and Amartya Sanyal. Language generation with replay: A learning-theoretic view of model collapse. *arXiv preprint arXiv:2603.11784*, 2026.

Vinod Raman, Jiaxun Li, and Ambuj Tewari. Generation through the lens of learning theory. In *Proceedings of the 38th Annual Conference on Learning Theory (COLT)*, volume 291 of *Proceedings of Machine Learning Research*, pp. 4740–4776, 2025. arXiv:2410.13714 (arXiv version lists authors as Li, Raman, Tewari).

Irene Strauss, Alexandra Butoi, and Ryan Cotterell. Generating in the limit with infinitely many hallucinations. *arXiv preprint arXiv:2606.28354*, 2026.

Alfred Tarski. Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften. I. *Monatshefte für Mathematik und Physik*, 37(1):361–404, 1930. doi: 10.1007/BF01696782.

Keith Wright. Identification of unions of languages drawn from an identifiable class. In *Proceedings of the Second Annual Workshop on Computational Learning Theory (COLT)*, pp. 328–333, 1989. doi: 10.1016/B978-0-08-094829-4.50026-X.

Yiming Zhang, Harshita Diddee, Susan Holm, Hanchen Liu, Xinyue Liu, Vinay Samuel, Barry Wang, and Daphne Ippolito. NoveltyBench: Evaluating language models for humanlike diversity. In *Conference on Language Modeling (COLM)*, 2025. arXiv:2504.05228.

Contents of the appendix

A Notation and the map of results **14**

A.1 Organization and conventions . . . 14

A.2 The map of results . . . 14

A.3 Notation . . . 14

B The principle and the matroid toolbox **16**

B.1 Closure spaces and matroids . . . 16

B.2 The matroid toolbox . . . 16

B.3 Escape or exhaustion . . . 17

C Uniform generation for every span **19**

C.1 Theorem 4.1: uniform generation for every span . . . 19

C.2 The two dimensions . . . 19

C.3 Corollary 4.2: three instances, and independent outputs . . . 20

C.4 Proposition 4.3: one language separates the theories . . . 20

D The other levels: counting only the examples, and own outputs **21**

D.1 Two constructions . . . 21

D.2 Proposition 5.1: counting only the examples . . . 21

D.3 Theorem 5.2: transfer with own outputs . . . 22

E Matroids: coordinate flats, the paraphrase trap, the hierarchy, and coincidence **23**

E.1 Coordinate flats . . . 23

E.2 The paraphrase trap . . . 24

E.3 Theorem 6.1: strict hierarchy . . . 25

E.4 Theorem 6.2: when the theories coincide . . . 25

F Locally Noetherian spans **26**

F.1 Theorem 5.3: the rank form . . . 27

F.2 Proposition 5.4: locally Noetherian spans . . . 27

F.3 The hierarchy of spans . . . 28

F.4 Explosion . . . 30

F.5 Master keys . . . 31

F.6 Same inside, opposite answers . . . 32

F.7 Safe bets and a region chosen after the first example . . . 33

F.8 Non-vacuity . . . 34

G Matroid intersection **34**

G.1 Proposition 7.1: capacity . . . 34

G.2 Proposition 7.2: rank in each matroid is not enough 35

H Extended related work **36**

H.1 Metric novelty and covering centres 36

H.2 Other work 36

A Notation and the map of results

A.1 Organization and conventions

How this appendix is organized. It follows the main text; the table lists where each result is stated and proved. Beyond the definitions of Section 2, the proofs use only the matroid toolbox of Appendix B, and that only where a matroid is involved.

result	stated in	proved in
Lemma 3.1 (escape or exhaustion)	Section 3	Appendix B
Theorem 4.1, Corollary 4.2, Proposition 4.3 (uniform generation for every span)	Section 4	Appendix C
Proposition 5.1, Theorem 5.2 (the other levels)	Section 5	Appendix D
Example 4.4, Theorems 6.1 and 6.2 (matroids)	Sections 4 and 6	Appendix E
Theorem 5.3, Propositions 5.4 and 5.6, Example 5.5 (locally Noetherian spans)	Section 5	Appendix F
Propositions 7.1 and 7.2 (several matroids)	Section 7	Appendix G

Conventions. $\mathrm{mindex}(Y)$ is the least-indexed element of a nonempty $Y \subseteq \mathcal{X}$ in the fixed enumeration $\xi_1, \xi_2, \ldots$. When a generator we define refers to an empty candidate set or to $\perp$, it outputs ξ_1 (the *totality convention*); the proofs show that this never happens at the rounds that matter. A generator recomputes its own earlier outputs from the prefix, so E_s and B_s are functions of $x_{1:s}$. We write $S_s = \{x_1, \ldots, x_s\}$, so that $c_s = |S_s| \leq s$. Thresholds are positive integers, and a uniform or non-uniform guarantee binds at every round t with $c_t \geq d^\star$ (Definition 2.3). Some results are stated and proved only here: the coordinate flats of Example E.1 (Appendix E); the hierarchy of locally Noetherian spans, the master keys, safe bets and a region chosen after the first example (Propositions F.1 to F.3 and Example F.4, Appendix F), to which the first open problem of Section 8 refers; and the example on covering centres in Appendix H.

Computer checks. Finite truncations of the following constructions (the closure spaces in the proof of Proposition F.1 and in Remark F.5, the matroids of Proposition 7.2, Theorem 6.1, and Examples 4.4 and E.1, the key-pair space of Example 5.5, the exhaustion argument of Lemma 3.1(c) on a rational grid, and random small matroids for the capacity bounds of Proposition 7.1) were checked by a short script included in the supplementary material. These are sanity checks of the infinite constructions, not proofs; the proofs below are self-contained.

A.2 The map of results

Figure 4 shows which results each proof uses. An arrow from one result to another means that the proof of the second cites the first; arrows implied by others are omitted, so a result also rests on everything above it along the arrows. Lemma 3.1 is used, directly or through Theorem 4.1, by every characterization; the toolbox is used wherever a matroid appears. Two citations run backwards between parts of results and are not drawn: the last clause of Proposition 5.1 uses Theorem 5.2, and the matroid sentence of Theorem 5.2 uses Corollary 4.2(ii); neither closes a cycle, since the parts they cite do not depend on them.

A.3 Notation

Table 2 collects the symbols of the paper, with the place where each is defined.

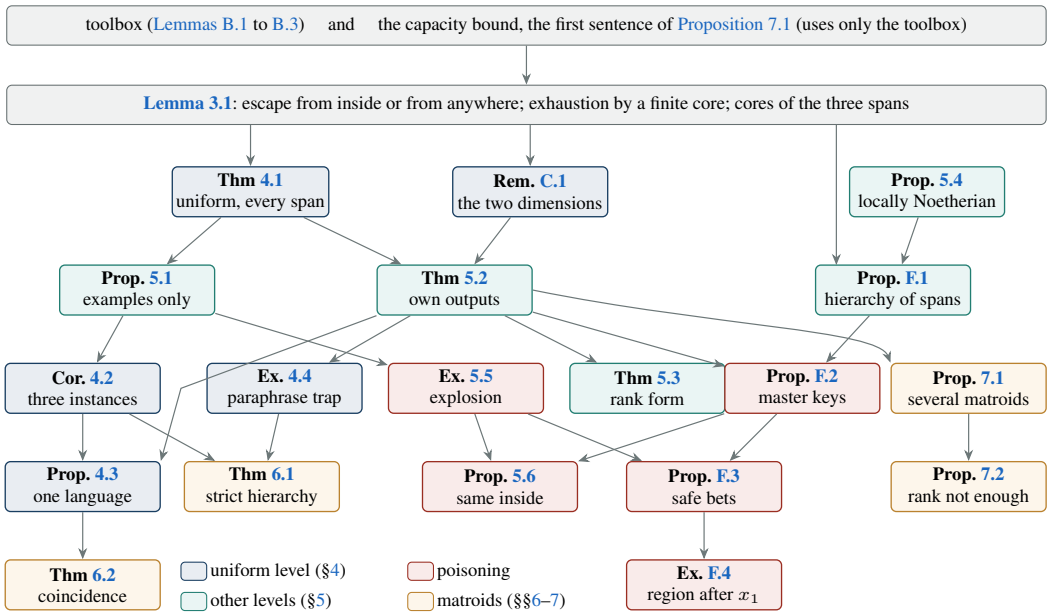

Figure 4: The map of results: an arrow means "is used in the proof of", up to transitivity. The two bars at the top are the foundations. The colours follow the structure of the paper: the uniform level for every span, the other levels, the second mechanism (poisoning, beyond locally Noetherian spans), and matroids.

Table 2: Notation.

symbol	meaning	defined in
$\mathcal{X} = \{\xi_1, \xi_2, \ldots\}$	countably infinite universe of strings, with a fixed enumeration	Section 2.1
$\mathcal{H}$, h, $\mathrm{supp}(h)$, K	class of languages, a language (as an indicator), its support, the target; every language is infinite	Section 2.1
$x_{1:s}$, S_s, c_s	the examples up to round s, their set, and their number	Section 2.1
$\hat{x}_s$, E_s	the output at round s; the history, examples and earlier outputs	Section 2.1
$\mathcal{H}(S)$, $\langle S \rangle_{\mathcal{H}}$	the languages consistent with S; the safe zone, the strings they all contain (what S entails)	Section 2.1
Ψ	a span: a monotone map with $E \subseteq \Psi(E)$, the strings that E makes redundant	Section 2.2
$B_s = \Psi(E_s)$, $B'_s = \Psi(S_s)$	the span of the full history; of the examples alone	Definition 2.1, Proposition 5.1
cl, cl_M, r_M	a closure operator; the closure and the rank of a matroid M	Example 2.2
finite core	a finite $C \subseteq A$ with $A \subseteq \Psi(C)$: A is spanned from inside	Section 2.3
finitely dominated	$A \subseteq \Psi(F)$ for a finite $F \subseteq \mathcal{X}$: spanned from anywhere	Section 2.3
locally Noetherian (FID)	every finitely dominated set has a finite core	Section 2.3
escape sequence; BEC	$o_i \notin \Psi(\{o_1, \ldots, o_{i-1}\})$ for all i; lengths bounded inside each finitely dominated set	Propositions 5.4 and F.1
Cdim, RCdim_M	closure dimension (Raman et al., 2025); its rank version for a matroid M	Section 2.3
$\mathrm{Coredim}_\Psi$, Domdim_Ψ	the core and the domination dimension of a class	Definition 2.4
$\rho_\cap$, $\mathrm{RCdim}_\cap$	common-independence capacity and dimension, for several matroids	Section 7
$d^\star$, t^*	a uniform or non-uniform threshold; the round from which a generator succeeds in the limit	Definition 2.3, Appendix D
Crit_t, n_t	the critical indices of the scaffold, and the largest of them	Appendix D
$\mathrm{mindex}(Y)$	the least-indexed element of Y in the fixed enumeration	Appendix A
p_j, $\neg p_j$ (or a_j, b_j)	literals; the key pairs $\{a_j, b_j\}$	Example 5.5

symbol	meaning	defined in
$(\mathcal{X}_0, \mathrm{cl}_0)$, a, b	the space of two master keys, which span everything together	Propositions F.1 and F.2
poisoned bet	a bet made for one language that, with examples of another, spans that other language	Section 5
safe region P	a set of bets that no finite history of examples and bets can poison	Proposition F.3

B THE PRINCIPLE AND THE MATROID TOOLBOX

B.1 CLOSURE SPACES AND MATROIDS

We number the axioms of a finitary closure space $(\mathcal{X}, \mathrm{cl})$ and of a finitary matroid (Example 2.2): (CL1) $A \subseteq \mathrm{cl}(A)$; (CL2) $A \subseteq A'$ implies $\mathrm{cl}(A) \subseteq \mathrm{cl}(A')$; (CL3) $\mathrm{cl}(\mathrm{cl}(A)) = \mathrm{cl}(A)$; (CL4) exchange, $y \in \mathrm{cl}(A \cup \{x\}) \setminus \mathrm{cl}(A)$ implies $x \in \mathrm{cl}(A \cup \{y\})$; (CL5) $\mathrm{cl}(A) = \bigcup\{\mathrm{cl}(F) : F \subseteq A \text{ finite}\}$. A span satisfies (CL1) and (CL2). In a matroid M a set I is *independent* if $x \notin \mathrm{cl}_M(I \setminus \{x\})$ for every $x \in I$, and a *basis* of A is a maximal independent subset of A. Since the rank is *defined* as the size of a smallest finite spanning set, "$r_M(A) = \infty$ if and only if A is not finitely dominated" holds by definition, without exchange.

B.2 THE MATROID TOOLBOX

In this subsection $M = (\mathcal{X}, \mathrm{cl}_M)$ is a finitary matroid. The two auxiliary lemmas come first; the toolbox lemma, which collects what the later proofs use, follows with its proof.

Lemma B.1 (Chains are independent)**.** *If $o_1, \dots, o_m \in \mathcal{X}$ satisfy $o_i \notin \mathrm{cl}_M(\{o_1, \dots, o_{i-1}\})$ for every $i \le m$, then the o_i are pairwise distinct and $\{o_1, \dots, o_m\}$ is M-independent.*

Proof. Distinctness. If $o_i = o_j$ with $j < i$ then $o_i \in \{o_1, \dots, o_{i-1}\} \subseteq \mathrm{cl}_M(\{o_1, \dots, o_{i-1}\})$ by (CL1).

Independence, by induction on j, with $I_j := \{o_1, \dots, o_j\}$. Base $j = 1$: $\{o_1\}$ is independent iff $o_1 \notin \mathrm{cl}_M(\emptyset)$, the $i = 1$ hypothesis.

Step. Assume I_{j-1} independent and $o_j \notin \mathrm{cl}_M(I_{j-1})$; we show $o_k \notin \mathrm{cl}_M(I_j \setminus \{o_k\})$ for all $k \le j$. For $k = j$ this is the hypothesis. For $k < j$ put $T := I_{j-1} \setminus \{o_k\}$, so $I_j \setminus \{o_k\} = T \cup \{o_j\}$. If $o_k \in \mathrm{cl}_M(T \cup \{o_j\})$, then since $o_k \notin \mathrm{cl}_M(T)$ by the inductive hypothesis, exchange (CL4) with $(A, x, y) = (T, o_j, o_k)$ gives $o_j \in \mathrm{cl}_M(T \cup \{o_k\}) = \mathrm{cl}_M(I_{j-1})$, a contradiction. □

Lemma B.2 (Steinitz bound)**.** *If I is a finite M-independent set and $I \subseteq \mathrm{cl}_M(F)$ for a finite F, then $|I| \le |F|$.*

Proof. Write $I = \{y_1, \dots, y_m\}$ and $|F| = n$. We prove by induction on $j \le \min(m, n)$ the statement P_j: *there is $F_j \subseteq F$ with $|F_j| = n - j$ and* $\mathrm{cl}_M(\{y_1, \dots, y_j\} \cup F_j) = \mathrm{cl}_M(F)$. Base: $F_0 = F$.

Step ($j < m$, $n - j \ge 1$). Write $F_j = \{f_1, \dots, f_{n-j}\}$, $G_0 = \{y_1, \dots, y_j\}$, $G_i = G_0 \cup \{f_1, \dots, f_i\}$. Then $y_{j+1} \in \mathrm{cl}_M(F) = \mathrm{cl}_M(G_{n-j})$ by P_j, while $y_{j+1} \notin \mathrm{cl}_M(G_0)$ by independence and (CL2). Let $i \ge 1$ be least with $y_{j+1} \in \mathrm{cl}_M(G_i)$; then $y_{j+1} \in \mathrm{cl}_M(G_{i-1} \cup \{f_i\}) \setminus \mathrm{cl}_M(G_{i-1})$, and (CL4) gives

$$f_i \in \mathrm{cl}_M(G_{i-1} \cup \{y_{j+1}\}).$$

Put $F_{j+1} := F_j \setminus \{f_i\}$. Every element of $\{y_1, \dots, y_j\} \cup F_j$ lies in $\mathrm{cl}_M(\{y_1, \dots, y_{j+1}\} \cup F_{j+1})$ (the element f_i by the last display, the others outright), so by (CL2)+(CL3) $\mathrm{cl}_M(F) \subseteq \mathrm{cl}_M(\{y_1, \dots, y_{j+1}\} \cup F_{j+1})$; the reverse inclusion holds because $\{y_1, \dots, y_{j+1}\} \cup F_{j+1} \subseteq \mathrm{cl}_M(F)$ and (CL3). This is P_{j+1}.

Conclusion. If $m > n$, run to $j = n$: $F_n = \emptyset$ and $y_{n+1} \in \mathrm{cl}_M(F) = \mathrm{cl}_M(\{y_1, \dots, y_n\})$, contradicting independence. Hence $m \le n$. □

Lemma B.3 (Matroid toolbox). *Let $M = (\mathcal{X}, \mathrm{cl}_M)$ be a finitary matroid.*

(C) (Chain bound; uses exchange.) *If $o_1, \dots, o_m \in A$ satisfy $o_i \notin \mathrm{cl}_M(D_i)$ for sets $D_i \supseteq \{o_1, \dots, o_{i-1}\}$, then $\{o_1, \dots, o_m\}$ is independent and $m \le r_M(A)$.*

(S) (Safe zone.) *For finite consistent $S \subseteq T$: $\langle S\rangle_{\mathcal{H}} \subseteq \langle T\rangle_{\mathcal{H}}$, and $\langle S\rangle_{\mathcal{H}} \subseteq \mathrm{supp}(h)$ for every $h \in \mathcal{H}(S)$.*

(T) (Closure trap.) *For finite consistent S: $S \subseteq \langle S\rangle_{\mathcal{H}}$ and $\mathcal{H}(\langle S\rangle_{\mathcal{H}}) = \mathcal{H}(S)$.*

(F) (Finite sets.) *For finite F and any finitary closure space, $\mathrm{cl}(F)$ is finitely dominated by F; in a matroid, $r_M(\mathrm{cl}_M(F)) = r_M(F) \le |F|$.*

(R) (Rank reconciliation; uses exchange.) *$r_M(A)$ is the supremum of the sizes of the finite independent subsets of A. If it is finite, every maximal independent subset I of A is finite and satisfies $A \subseteq \mathrm{cl}_M(I)$.*

Proof. (C) Since $\{o_1, \dots, o_{i-1}\} \subseteq D_i$, (CL2) gives $\mathrm{cl}_M(\{o_1, \dots, o_{i-1}\}) \subseteq \mathrm{cl}_M(D_i)$, so $o_i \notin \mathrm{cl}_M(\{o_1, \dots, o_{i-1}\})$ for every i, and Lemma B.1 gives independence. If $r_M(A) < \infty$ then $A \subseteq \mathrm{cl}_M(F)$ with $|F| = r_M(A)$, and Lemma B.2 gives $m \le |F|$.

(S) $\langle S\rangle_{\mathcal{H}} = \bigcap_{h' \in \mathcal{H}(S)} \mathrm{supp}(h') \subseteq \mathrm{supp}(h)$ for $h \in \mathcal{H}(S)$. For consistent $S \subseteq T$, $\emptyset \ne \mathcal{H}(T) \subseteq \mathcal{H}(S)$, and intersecting over the smaller family gives the larger set.

(T) $S \subseteq \langle S\rangle_{\mathcal{H}}$ since every $h \in \mathcal{H}(S)$ contains S. If $h \in \mathcal{H}(S)$ then $\langle S\rangle_{\mathcal{H}} \subseteq \mathrm{supp}(h)$ by (S), so $h \in \mathcal{H}(\langle S\rangle_{\mathcal{H}})$; conversely $S \subseteq \langle S\rangle_{\mathcal{H}} \subseteq \mathrm{supp}(h)$ gives $h \in \mathcal{H}(S)$.

(F) $\mathrm{cl}(F)$ is dominated by F. In a matroid $r_M(F) \le |F|$ by (CL1), and $r_M(\mathrm{cl}_M(F)) = r_M(F)$ because the dominating sets coincide: $F \subseteq \mathrm{cl}_M(F')$ implies $\mathrm{cl}_M(F) \subseteq \mathrm{cl}_M(F')$ by (CL2)+(CL3), and $F \subseteq \mathrm{cl}_M(F)$ by (CL1).

(R), the supremum is at most the rank. Let $\sigma(A)$ be the supremum of the sizes of finite independent subsets of A. A finite independent $I \subseteq A \subseteq \mathrm{cl}_M(F)$ with $|F| = r_M(A)$ has $|I| \le |F|$ by Lemma B.2, so $\sigma \le r_M$.

(R), the exchange step. Now let $I \subseteq A$ be independent and suppose $a \in A \setminus \mathrm{cl}_M(I)$. Then $I \cup \{a\}$ is independent: $a \notin \mathrm{cl}_M(I)$ directly, and for $x \in I$, if $x \in \mathrm{cl}_M((I \setminus \{x\}) \cup \{a\})$ then, as $x \notin \mathrm{cl}_M(I \setminus \{x\})$, (CL4) gives $a \in \mathrm{cl}_M(I)$, a contradiction.

(R), the rank is at most the supremum. If $\sigma(A) = k < \infty$, the supremum is attained by some independent $I \subseteq A$ with $|I| = k$, no $a \in A \setminus \mathrm{cl}_M(I)$ can exist, and so $A \subseteq \mathrm{cl}_M(I)$ and $r_M(A) \le k$.

(R), maximal independent sets. Finally, if $r_M(A) < \infty$, a maximal independent $I \subseteq A$ has all finite subsets independent and of size at most $r_M(A)$, so I is finite; by maximality no $a \in A \setminus \mathrm{cl}_M(I)$ exists, so $A \subseteq \mathrm{cl}_M(I)$. □

B.3 Escape or exhaustion

Lemma 3.1 (Escape or exhaustion). *Let Ψ be a span, or any monotone map on subsets of $\mathcal{X}$, and let $B_s = \Psi(E_s)$.*

(a) (Showing the safe zone reveals nothing.) *If S and T are finite, $\mathcal{H}(S) \ne \emptyset$, and $S \subseteq T \subseteq \langle S\rangle_{\mathcal{H}}$, then $\mathcal{H}(T) = \mathcal{H}(S)$ and $\langle T\rangle_{\mathcal{H}} = \langle S\rangle_{\mathcal{H}}$.*

(b) (Escape.) *If A has no finite core, then $A \not\subseteq \Psi(E)$ for every finite $E \subseteq A$* (from inside). *If A is not finitely dominated, then $A \not\subseteq \Psi(E)$ for every finite E* (from anywhere).

(c) (Exhaustion.) *Let S be finite with $\mathcal{H}(S) \ne \emptyset$, and suppose $A = \langle S\rangle_{\mathcal{H}}$ has a finite core C. Then no deterministic generator achieves a uniform threshold $d^\star \le |S|$, nor non-uniform thresholds with $d^\star(h) \le |S|$ for every $h \in \mathcal{H}(S)$. Likewise, if a language $K = \mathrm{supp}(h)$ has a finite core, then on every complete enumeration of K no round succeeds once the core has appeared. Both statements hold even for the smaller span $\Psi(\{x_1, \dots, x_s\})$ of the examples alone, which for Ψ the identity is plain novelty.*

(d) (Cores.) *For the identity, a set has a finite core iff it is finite; for a finitary matroid, iff it has finite rank, and then every basis of the set is a finite core; for $\Psi = \bigcup_{i \le k} \mathrm{cl}_{M_i}$, iff it contains no infinite set that is independent in every M_i. Each of these three spans is locally Noetherian.*

Proof. (a) If $h \in \mathcal{H}(S)$ then $T \subseteq \langle S \rangle_{\mathcal{H}} \subseteq \mathrm{supp}(h)$ by Lemma B.3(S), so $h \in \mathcal{H}(T)$; conversely $S \subseteq T$ gives $\mathcal{H}(T) \subseteq \mathcal{H}(S)$. The two version spaces agree, and so do their intersections: $\langle T \rangle_{\mathcal{H}} = \langle S \rangle_{\mathcal{H}}$.

(b) Both halves restate the definitions. If $E \subseteq A$ is finite and $A \subseteq \Psi(E)$, then E is a finite core of A; if $A \subseteq \Psi(E)$ for some finite $E \subseteq \mathcal{X}$, then A is finitely dominated. In the games the history E_s is finite, so each half applies to $B_s = \Psi(E_s)$, and to the smaller $\Psi(S_s)$, whenever its hypothesis on the set spanned holds.

(c), uniform and non-uniform part. Write $S = \{s_1, \ldots, s_p\}$ and $C = \{c_1, \ldots, c_q\}$ (the s_i pairwise distinct, the c_j pairwise distinct, $S \cap C$ arbitrary; $q = 0$ allowed), let

$$\pi = (s_1, \ldots, s_p, c_1, \ldots, c_q), \qquad n = p + q, \qquad \text{and} \qquad T = S \cup C.$$

The version space does not move. $S \subseteq A$ by Lemma B.3(T) and $C \subseteq A$ by hypothesis, so $S \subseteq T \subseteq A$, and (a) gives $\mathcal{H}(T) = \mathcal{H}(S)$ and $\langle T \rangle_{\mathcal{H}} = A$; every prefix of π is legal for every $h \in \mathcal{H}(S)$.

The threshold is reached. The first p entries of π are distinct, so $c_{d^\star} = d^\star$ at round $d^\star \leq p \leq n$; the guarantee binds at every round $\geq d^\star$, in particular at round n, for any committed pair whose stream begins with π. In the non-uniform game, whichever $h \in \mathcal{H}(S)$ is committed below has $d^\star(h) \leq p$, so its guarantee binds at round n as well.

Evaluate. Let $o = G(\pi)$; by determinism $\hat{x}_n = o$ on every stream beginning with π.

Case I: $o \in A$. Then

$$o \in \Psi(C) \subseteq \Psi(T) \subseteq \Psi(T \cup \{\hat{x}_{<n}\}) = B_n$$

by monotonicity, as $C \subseteq T$; the middle term is the span of the examples alone, which gives the examples-only form. Commit any $h^* \in \mathcal{H}(S)$ and the stream $\sigma = (\pi, s_1, s_1, \ldots)$, legal since all its entries lie in $T \subseteq A \subseteq \mathrm{supp}(h^*)$: novelty fails at round n.

Case II: $o \notin A = \bigcap_{h \in \mathcal{H}(S)} \mathrm{supp}(h)$. Some $h' \in \mathcal{H}(S)$ has $o \notin \mathrm{supp}(h')$; commit (h', σ), legal for h': validity fails at round n.

Conclusion. In both cases the adversary committed a fixed pair after inspecting the fixed map G on one finite prefix, so the counterexample is oblivious. A generator achieving a threshold $d^\star \leq p$ would succeed at round n on this pair; none does.

(c), limit part. Let C be a finite core of $K = \mathrm{supp}(h)$, and fix a complete enumeration of K. Each $c \in C$ appears in the enumeration at some finite round; let τ be the largest such round ($\tau = 1$ if $C = \emptyset$). For $t \geq \tau$, $C \subseteq \{x_1, \ldots, x_t\} \subseteq E_t$, so

$$K \subseteq \Psi(C) \subseteq \Psi(\{x_1, \ldots, x_t\}) \subseteq \Psi(E_t) = B_t$$

by monotonicity, and $\mathrm{supp}(h) \setminus B_t = \emptyset$, already for the examples-only span.

(d), the identity. If $C \subseteq A$ is finite and $A \subseteq \Psi(C) = C$, then $A = C$ is finite, and a finite set is its own core. If $A \subseteq F$ for a finite F, then A is finite and so has a finite core: the identity is locally Noetherian.

(d), a finitary matroid M. If A has a finite core C, then $A \subseteq \mathrm{cl}_M(C)$ and $r_M(A) \leq |C| < \infty$ by the definition of rank.

Finite rank gives a finite core. Conversely, let $r_M(A) < \infty$. By Lemma B.3(R) the finite independent subsets of A have at most $r_M(A)$ elements, so one of maximum size exists; it is a basis, since a strictly larger independent set would contain a larger finite one (subsets of independent sets are independent, by (CL2)). Every basis I of A is finite, as its finite subsets are independent and so have at most $r_M(A)$ elements, and it spans A: if some $a \in A$ had $a \notin \mathrm{cl}_M(I)$, then $I \cup \{a\}$ would be independent by the exchange step in the proof of Lemma B.3(R) ($a \notin \mathrm{cl}_M(I)$; and if $x \in \mathrm{cl}_M((I \setminus \{x\}) \cup \{a\})$ for some $x \in I$, then, as $x \notin \mathrm{cl}_M(I \setminus \{x\})$, (CL4) would give $a \in \mathrm{cl}_M(I)$), contradicting maximality. So every basis of A is a finite core.

Local Noetherianity. Finally, if $A \subseteq \mathrm{cl}_M(F)$ for a finite F, then $r_M(A) \leq |F| < \infty$, and A has a finite core: every finitary matroid is locally Noetherian.

(d), several matroids. For $\Psi = \bigcup_{i \leq k} \mathrm{cl}_{M_i}$, Proposition 7.1, whose proof in Appendix G uses only the toolbox, shows that A has a finite core iff it is finitely dominated iff it contains no infinite set that is independent in every M_i; in particular this span is locally Noetherian. □

C UNIFORM GENERATION FOR EVERY SPAN

C.1 THEOREM 4.1: UNIFORM GENERATION FOR EVERY SPAN

Theorem 4.1 (Uniform generation for every span)**.** *A class $\mathcal{H}$ of any cardinality is uniformly generatable if and only if* $\mathrm{Coredim}_\Psi(\mathcal{H}) < \infty$*, and then exactly* $\mathrm{Coredim}_\Psi(\mathcal{H})+1$ *distinct examples are needed: this threshold is achievable, and no smaller one is. In particular, a generator that is valid and novel at every round exists if and only if* $\mathrm{Coredim}_\Psi(\mathcal{H}) = 0$*. Moreover* $\mathrm{Coredim}_\Psi(\mathcal{H}) \in \{\mathrm{Cdim}(\mathcal{H}), \infty\}$.

Proof. Write $Z(S) = \langle S \rangle_{\mathcal{H}}$ for finite consistent S; safe zones contain S and grow with S (Lemma B.3(S),(T)). A *core witness* is a consistent set of distinct strings whose safe zone has a finite core. We use only that $\mathrm{Coredim}_\Psi(\mathcal{H})$ is the supremum of the sizes of core witnesses, not that subsets of witnesses are witnesses.

Lower bound. Let $d^\star \geq 1$ with $d^\star \leq \mathrm{Coredim}_\Psi(\mathcal{H})$. There is a core witness S with $|S| \geq d^\star$: if the dimension is infinite, witnesses of unbounded size exist, and if it is a finite $D \geq d^\star$, the supremum is attained by a witness of size D. Lemma 3.1(c) shows that no deterministic generator achieves threshold $d^\star$.

Generator. Let $D = \mathrm{Coredim}_\Psi(\mathcal{H}) < \infty$ and $Z_t = Z(S_t)$. Output $\mathrm{mindex}(Z_t \setminus \Psi(E_t))$ if this set is nonempty and $\mathrm{mindex}(Z_t)$ otherwise (totality convention if S_t is inconsistent).

The history stays in the safe zone. Fix h and a stream of $\mathrm{supp}(h)$, so that $h \in \mathcal{H}(S_t)$ for every t. Then $E_t \subseteq Z_t$: $S_t \subseteq Z_t$, and each earlier output $\hat{x}_s$ lies in $Z_s \subseteq Z_t$.

Escape. If $c_t \geq D+1$, then Z_t has no finite core, since S_t would otherwise be a core witness of size $c_t > D$; so $Z_t \not\subseteq \Psi(E_t)$ for the finite $E_t \subseteq Z_t$ (Lemma 3.1(b), from inside), and the output lies in

$$Z_t \setminus B_t \subseteq \mathrm{supp}(h) \setminus B_t.$$

So threshold $D+1$ is achieved, and by the lower bound no smaller one is; if $\mathrm{Coredim}_\Psi(\mathcal{H}) = \infty$, the lower bound excludes every threshold.

Valid and novel at every round. Every stream has $c_1 = 1$, so a generator succeeds at every round on every stream iff it achieves threshold 1. By the above, the generator achieves threshold 1 when $\mathrm{Coredim}_\Psi(\mathcal{H}) = 0$, and the lower bound excludes threshold 1 when $\mathrm{Coredim}_\Psi(\mathcal{H}) \geq 1$, including the infinite case. So such a generator exists iff $\mathrm{Coredim}_\Psi(\mathcal{H}) = 0$.

Free or fatal. A span is extensive, so a finite safe zone is its own core and every Cdim witness is a core witness: $\mathrm{Cdim} \leq \mathrm{Coredim}_\Psi$. If some core witness S has an infinite safe zone A, then for every $n \geq |S|$ pick T with $S \subseteq T \subseteq A$ and $|T| = n$; by Lemma 3.1(a), $\mathcal{H}(T) = \mathcal{H}(S)$ and $Z(T) = A$, which keeps its finite core, so T is a core witness of size n and $\mathrm{Coredim}_\Psi(\mathcal{H}) = \infty$. Otherwise every core witness has a finite safe zone and is a Cdim witness, so $\mathrm{Coredim}_\Psi \leq \mathrm{Cdim}$. Hence

$$\mathrm{Coredim}_\Psi(\mathcal{H}) \in \{\mathrm{Cdim}(\mathcal{H}), \infty\}.$$

Hypotheses used. Only monotonicity of Ψ entered the lower bound, and no property of Ψ entered the generator; extensivity is used only for free or fatal. □

C.2 THE TWO DIMENSIONS

Remark C.1 (The two dimensions). A finite core is a finite dominating set, so every core witness is a domination witness and $\mathrm{Coredim}_\Psi(\mathcal{H}) \leq \mathrm{Domdim}_\Psi(\mathcal{H})$ for every span and class. If Ψ is locally Noetherian, the two witness conditions coincide, and $\mathrm{Coredim}_\Psi = \mathrm{Domdim}_\Psi$ for every class. For a finitary closure space the converse holds as well: if A is finitely dominated but has no finite core, then A is infinite (a finite set is its own core), and the one-language class $\{\mathbf{1}_A\}$ has $\mathrm{Domdim}_{\mathrm{cl}} = \infty$ (proof of Theorem 5.3) but $\mathrm{Coredim}_{\mathrm{cl}} = 0$, since every consistent set has safe zone A, which has no finite core. For the identity both dimensions are Cdim, and for a finitary matroid both are RCdim_M: by Lemma 3.1(d) the conditions "has a finite core" and "is finitely dominated" both read "is finite", respectively "has finite rank".

C.3 Corollary 4.2: three instances, and independent outputs

Corollary 4.2 (Three instances)**.** (i) *For the identity, uniform generation, plain or with repeats also forbidden, is possible if and only if* $\mathrm{Cdim}(\mathcal{H}) < \infty$ *(Raman et al., 2025), with optimal threshold* $\mathrm{Cdim}(\mathcal{H}) + 1$*; in particular a generator that is valid and unseen at every round exists if and only if* $\mathrm{Cdim}(\mathcal{H}) = 0$*.* (ii) *For a finitary matroid* M *the same holds with the rank-closure dimension* RCdim_M*, and* $\mathrm{RCdim}_M(\mathcal{H}) \in \{\mathrm{Cdim}(\mathcal{H}), \infty\}$*.* (iii) *For* $\Psi = \bigcup_{i \le k} \mathrm{cl}_{M_i}$ *the same holds with the common-independence dimension of Section 7.*

Proof. The three core dimensions. By Lemma 3.1(d), the witness condition of $\mathrm{Coredim}_\Psi$, that the safe zone has a finite core, reads "the safe zone is finite" for the identity, "it has finite rank" for a finitary matroid, and "it contains no infinite common independent set", that is, $\rho_\cap < \infty$ (Proposition 7.1), for $\Psi = \bigcup_{i \le k} \mathrm{cl}_{M_i}$. So $\mathrm{Coredim}_\Psi$ is Cdim, RCdim_M and $\mathrm{RCdim}_\cap$, respectively, and Theorem 4.1 gives each characterization, the optimal threshold and the criterion for success at every round.

Free or fatal. Each of the three maps is extensive, so free or fatal gives $\mathrm{RCdim}_M(\mathcal{H}) \in \{\mathrm{Cdim}(\mathcal{H}), \infty\}$, and likewise for $\mathrm{RCdim}_\cap$.

The identity: repeats and plain novelty. For the identity, Theorem 4.1 concerns the history $B_s = E_s$, which forbids repeats as well. Plain novelty is the identity with the examples-only span $B_s = S_s$, for which Proposition 5.1 gives the same characterization and the same optimal threshold $\mathrm{Cdim}(\mathcal{H}) + 1$; as in the proof of Theorem 4.1, a generator that is valid and unseen at every round is exactly one that achieves threshold 1, which is possible iff $\mathrm{Cdim}(\mathcal{H}) = 0$. □

Remark C.2 (Independent outputs). Let M be a finitary matroid, and suppose a generator succeeds at rounds $s_1 < s_2 < \cdots$. Each $\hat{x}_{s_j}$ lies outside $B_{s_j} \supseteq \mathrm{cl}_M(\{\hat{x}_{s_1}, \ldots, \hat{x}_{s_{j-1}}\})$, so by Lemma B.3(C) the outputs at successful rounds form an independent set; if the generator succeeds from round one, its entire output sequence is independent. Before a guarantee takes effect, outputs may repeat or be dependent (the generator in the proof of Theorem 4.1 falls back to $\mathrm{mindex}(Z_t)$ when no candidate is novel), so the independent set is the *successful tail*, not the full sequence. This is the design content of including own outputs in the history; the lower bounds do not use it.

C.4 Proposition 4.3: one language separates the theories

Proposition 4.3 (One language separates the theories)**.** (i) *For every span, every* $\mathcal{H}$ *and each game, novel generatability implies plain generatability, with the same generator and thresholds.* (ii) *The converse fails already for one language: for the linear matroid on* $\mathcal{X} = \mathbb{Q}^2$ *and* $\mathcal{H} = \{\mathbf{1}_{\mathbb{Q}^2}\}$*, we have* $\mathrm{Cdim}(\mathcal{H}) = 0$*, so* $\mathcal{H}$ *is plainly generatable from the first example, while* $\mathrm{RCdim}_M(\mathcal{H}) = \infty$ *and* $\mathcal{H}$ *is not rank-novel generatable even in the limit.*

Proof. (i) A span is extensive, so $B_s = \Psi(E_s) \supseteq E_s \supseteq S_s$ and $\mathrm{supp}(h) \setminus B_s \subseteq \mathrm{supp}(h) \setminus S_s$: novel success implies plain success pointwise, and every generator and threshold transfers unchanged in each game.

(ii), the matroid and the class. $\mathcal{X} = \mathbb{Q}^2$ is countable, and $\mathrm{cl}_M = \mathrm{span}_\mathbb{Q}$ is a finitary matroid (exchange: $y \in \mathrm{span}(A \cup \{x\}) \setminus \mathrm{span}(A)$ forces a nonzero x-coefficient; finitary: linear combinations are finite), with $r_M(\mathbb{Q}^2) = 2$. The class $\mathcal{H} = \{\mathbf{1}_{\mathbb{Q}^2}\}$ is a singleton, and every set is consistent with safe zone $\mathbb{Q}^2$.

Plain side. Every safe zone is infinite, so $\mathrm{Cdim}(\mathcal{H}) = 0$, and "output the least-indexed unseen point" is valid and unseen from round one.

Rank side. Any d distinct points are a witness (safe zone $\mathbb{Q}^2$, of rank 2), so $\mathrm{RCdim}_M(\mathcal{H}) = \infty$, and $\mathcal{H}$ is not uniformly generatable (Corollary 4.2(ii)).

Rank side, in the limit. Not even in the limit: a basis of $\mathbb{Q}^2$ is a finite core of the only language (Lemma 3.1(d)), so the exhaustion side of Theorem 5.2(a) applies; concretely, once a basis of $\mathbb{Q}^2$ has been enumerated the history spans the plane. □

D THE OTHER LEVELS: COUNTING ONLY THE EXAMPLES, AND OWN OUTPUTS

Both results of Section 5 are proved with two constructions, each used once for each span. Write $B'_s = \Psi(S_s)$ for the examples-only span and $B_s = \Psi(E_s)$ for the full one. Since $S_s \subseteq E_s$ and Ψ is monotone,

$$B'_s \subseteq B_s:$$

a generator that succeeds against B at a round succeeds against B' there.

D.1 TWO CONSTRUCTIONS

The critical-language scaffold (relabelled from Kleinberg & Mullainathan, 2024). Fix an injective enumeration $h_1, h_2, \ldots$ of a countable $\mathcal{H}$ and write $L_n = \mathrm{supp}(h_n)$; for finite $\mathcal{H}$ of size N read "$n \le t$" as "$n \le \min(t, N)$". The adversary commits $h = h_z$ and a complete enumeration (x_t) of $K := L_z$. Say L_n is *critical at* t if (a) $S_t \subseteq L_n$ and (b) for every $i \le n$ with $S_t \subseteq L_i$ we have $L_n \subseteq L_i$; let

$$\mathrm{Crit}_t = \{n \le t : L_n \text{ critical at } t\},$$

a function of S_t and the enumeration of $\mathcal{H}$ alone.

(K1) Eventual criticality of the target. $\mathrm{Bad} := \{i \le z : K \not\subseteq L_i\}$ is finite; for $i \in \mathrm{Bad}$ fix $v_i \in K \setminus L_i$, which appears by some round τ_i; let $t^+ = \max_i \tau_i$ (1 if $\mathrm{Bad} = \emptyset$). For $t \ge t^+$, (a) holds for $n = z$, and any $i \le z$ with $S_t \subseteq L_i$ is not in Bad, i.e. $K \subseteq L_i$, which is (b).

(K2) Chain property. If $i < j$ are both critical at t then $L_j \subseteq L_i$: (a) for L_i gives $S_t \subseteq L_i$, and (b) for L_j at index i applies.

(K3) Convergence. For $t \ge t^*$, where

$$t^* := \max(z, t^+),$$

we have $z \in \mathrm{Crit}_t$, so

$$n_t := \max \mathrm{Crit}_t \ge z$$

exists and $S_t \subseteq L_{n_t} \subseteq K$, by (a) and, if $n_t \ne z$, by (K2). This holds even if n_t changes indefinitely; no identification is assumed, and no property of the span is used.

Level selection. Let $\mathcal{H} = \bigcup_n \mathcal{H}_n$ for a chain $\mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq \cdots$ and let $d_1, d_2, \ldots$ be positive integers. On a prefix $x_{1:s}$ let

$$N_s = \{n \le s : d_n \le c_s\};$$

if $N_s \ne \emptyset$ let

$$\tilde{n}_s = \max N_s \qquad \text{and} \qquad Z_s = \langle S_s \rangle_{\mathcal{H}_{\tilde{n}_s}} \quad (\perp \text{ if } \mathcal{H}_{\tilde{n}_s}(S_s) = \emptyset).$$

For $h \in \mathcal{H}$ let

$$m_h = \min\{n : h \in \mathcal{H}_n\} \qquad \text{and} \qquad d^\star(h) = \max(d_{m_h}, m_h),$$

which depends only on h and the fixed chain and sequence.

Fact. On a stream of $\mathrm{supp}(h)$, every round s with $c_s \ge d^\star(h)$ has $m_h \le c_s \le s$ and $d_{m_h} \le c_s$, so $m_h \in N_s$ and $\tilde{n}_s \ge m_h$; hence $h \in \mathcal{H}_{m_h} \subseteq \mathcal{H}_{\tilde{n}_s}$ and $h \in \mathcal{H}_{\tilde{n}_s}(S_s)$, so

$$S_s \subseteq Z_s \subseteq \mathrm{supp}(h)$$

(Lemma B.3(S),(T) for the class $\mathcal{H}_{\tilde{n}_s}$); and $c_s \ge d_{\tilde{n}_s}$, as $\tilde{n}_s \in N_s$.

The generator. Given a forbidden set Φ_s computed from the prefix, the *level-selection generator* outputs $\mathrm{mindex}(Z_s \setminus \Phi_s)$ when $N_s \ne \emptyset$, $Z_s \ne \perp$ and this set is nonempty, and ξ_1 otherwise.

D.2 PROPOSITION 5.1: COUNTING ONLY THE EXAMPLES

Proposition 5.1 (Counting only the examples). *Let Ψ be a span and measure novelty against the examples alone, $B_s = \Psi(\{x_1, \ldots, x_s\})$. Then $\mathcal{H}$ is uniformly generatable iff $\mathrm{Coredim}_\Psi(\mathcal{H}) < \infty$, with optimal threshold $\mathrm{Coredim}_\Psi(\mathcal{H}) + 1$; non-uniformly iff $\mathcal{H} = \bigcup_n \mathcal{H}_n$ for a chain $\mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq$*

$\cdots$ with $\mathrm{Coredim}_\Psi(\mathcal{H}_n) < \infty$ for all n; and, if countable, in the limit iff no language has a finite core. With the generator's own outputs counted, $B_s = \Psi(E_s)$, the uniform answer is the same for every span, and the other two are the same (the limit for countable classes) whenever the span is locally Noetherian.

Proof. Uniform. The generator of Theorem 4.1 achieves threshold $\mathrm{Coredim}_\Psi(\mathcal{H}) + 1$ against B, hence against B'. Its lower bound is Lemma 3.1(c), which holds for the examples-only span: no threshold $d^\star \le \mathrm{Coredim}_\Psi(\mathcal{H})$ is achievable against B', hence none against B, and if the dimension is infinite, no threshold is achievable against either. So for B', as for B, $\mathcal{H}$ is uniformly generatable iff $\mathrm{Coredim}_\Psi(\mathcal{H}) < \infty$, with optimal threshold $\mathrm{Coredim}_\Psi(\mathcal{H}) + 1$; the uniform answer is the same for every span.

Non-uniform, necessity. Let G achieve the non-uniform guarantee against B', let d_h be the least valid threshold for h, and put $\mathcal{H}_n = \{h : d_h \le n\}$, a chain with union $\mathcal{H}$. Then G achieves the uniform threshold n for $\mathcal{H}_n$ against B': on a stream of $\mathrm{supp}(h)$ with $h \in \mathcal{H}_n$, every round with $c_t \ge n$ has $c_t \ge d_h$, where the guarantee for h binds. By the uniform part, $\mathrm{Coredim}_\Psi(\mathcal{H}_n) \le n - 1$ (with $\mathrm{Coredim}_\Psi(\emptyset) = 0$).

Non-uniform, sufficiency. Given a chain with $\mathrm{Coredim}_\Psi(\mathcal{H}_n) < \infty$ for all n, run level selection with

$$d_n = \mathrm{Coredim}_\Psi(\mathcal{H}_n) + 1 \qquad \text{and} \qquad \Phi_s = \Psi(S_s).$$

At a round with $c_s \ge d^\star(h)$ on a stream of $\mathrm{supp}(h)$, the Fact gives $S_s \subseteq Z_s \subseteq \mathrm{supp}(h)$ and

$$c_s \ge d_{\tilde{n}_s} > \mathrm{Coredim}_\Psi(\mathcal{H}_{\tilde{n}_s}),$$

so the consistent set S_s of c_s distinct points is not a core witness for $\mathcal{H}_{\tilde{n}_s}$, and Z_s has no finite core. As S_s is a finite subset of Z_s, $Z_s \not\subseteq \Psi(S_s)$ (Lemma 3.1(b), from inside), and the output lies in $Z_s \setminus B'_s \subseteq \mathrm{supp}(h) \setminus B'_s$. So the non-uniform guarantee holds with the thresholds $d^\star(h)$.

Countable classes. If every finite subclass of $\mathcal{H} = \{h_1, h_2, \ldots\}$ has finite core dimension, the chain $\mathcal{H}_n = \{h_1, \ldots, h_n\}$ qualifies. Conversely, necessity gives a chain, and a finite $\mathcal{H}' \subseteq \mathcal{H}$ lies in some $\mathcal{H}_N$. The core dimension does not increase on passing to a subclass: the generator of Theorem 4.1 for $\mathcal{H}_N$ at threshold $\mathrm{Coredim}_\Psi(\mathcal{H}_N) + 1$ is one for $\mathcal{H}'$, so $\mathrm{Coredim}_\Psi(\mathcal{H}') \le \mathrm{Coredim}_\Psi(\mathcal{H}_N) < \infty$ by the lower bound of Theorem 4.1.

Limit. Let $\mathcal{H}$ be countable. If no language has a finite core, output $\mathrm{mindex}(L_{n_t} \setminus \Psi(S_t))$ when $\mathrm{Crit}_t \ne \emptyset$ and this set is nonempty, and ξ_1 otherwise. For $t \ge t^*$ the scaffold gives $S_t \subseteq L_{n_t} \subseteq K$; the language L_{n_t} has no finite core and S_t is a finite subset of it, so $L_{n_t} \not\subseteq \Psi(S_t)$ (Lemma 3.1(b), from inside), and the output lies in $L_{n_t} \setminus B'_t \subseteq K \setminus B'_t$. If some language has a finite core, Lemma 3.1(c), whose limit part holds for the examples-only span, shows that $\mathcal{H}$ is not generatable in the limit.

Own outputs counted. The uniform answer was settled above. If Ψ is locally Noetherian, Theorem 5.2, whose proof below does not use this proposition, characterizes non-uniform generatability against B by the same chain condition and, for countable classes, generatability in the limit against B by the same condition, that no language has a finite core.

The identity. For Ψ the identity, $B'_s = S_s$ is plain novelty, and a set has a finite core iff it is finite (Lemma 3.1(d)), so $\mathrm{Coredim} = \mathrm{Cdim}$. Languages are infinite, so none has a finite core, and every countable class is generatable in the limit (Kleinberg & Mullainathan, 2024). A finite class $\mathcal{H}'$ has finite Cdim: a witness S lies inside its own finite safe zone (Lemma B.3(T)), which is the intersection of the supports of a nonempty subfamily of $\mathcal{H}'$, and there are finitely many such intersections. So the countable form applies, and every countable class is non-uniformly generatable (Raman et al., 2025). □

D.3 Theorem 5.2: Transfer with own outputs

Theorem 5.2 (Transfer with own outputs)**.** *Let Ψ be a span and $B_s = \Psi(E_s)$.*

(a) (Limit.) *Let $\mathcal{H}$ be countable. If no language of $\mathcal{H}$ is finitely dominated, $\mathcal{H}$ is generatable in the limit; if some language has a finite core, it is not. So for a locally Noetherian span, $\mathcal{H}$ is generatable in the limit iff no language has a finite core.*

(b) (Non-uniform.) *If $\mathcal{H}$ is non-uniformly generatable, then $\mathcal{H} = \bigcup_n \mathcal{H}_n$ for a chain with $\mathrm{Coredim}_\Psi(\mathcal{H}_n) < \infty$ for all n; conversely, such a chain with $\mathrm{Domdim}_\Psi(\mathcal{H}_n) < \infty$ for all n*

suffices. So for a locally Noetherian span, $\mathcal{H}$ is non-uniformly generatable iff such a chain exists, and a countable $\mathcal{H}$ is iff each of its finite subclasses has finite core dimension.

For matroids, a countable class is generatable in the limit iff every language has infinite rank.

Proof. (a), escape side. Assume no language of $\mathcal{H}$ is finitely dominated. Output $\mathrm{mindex}(L_{n_t} \setminus \Psi(E_t))$ when $\mathrm{Crit}_t \neq \emptyset$ and this set is nonempty, and ξ_1 otherwise. Unlike the set-function generator of Kleinberg & Mullainathan (2024), this generator depends on the prefix also through its own outputs, which is licit. For $t \geq t^*$ the scaffold gives $L_{n_t} \subseteq K$; the language L_{n_t} is not finitely dominated and E_t is finite, so $L_{n_t} \not\subseteq \Psi(E_t)$ (Lemma 3.1(b), from anywhere), and the output lies in

$$L_{n_t} \setminus B_t \subseteq K \setminus B_t.$$

The history may contain earlier outputs outside L_{n_t}, chosen while another language was selected; escape from anywhere does not need the history inside L_{n_t}.

(a), exhaustion side. If some language has a finite core, Lemma 3.1(c) shows that on every complete enumeration of it no round succeeds once the core has appeared.

(a), locally Noetherian spans. A finite core makes a set finitely dominated, and under local Noetherianity conversely; so "no language is finitely dominated" and "no language has a finite core" coincide, and the two sides combine.

(b), necessity. Let G achieve the non-uniform guarantee against B, and define d_h and $\mathcal{H}_n = \{h : d_h \leq n\}$ as in the proof of Proposition 5.1. Then G achieves the uniform threshold n for $\mathcal{H}_n$ against B, and the lower bound of Theorem 4.1 gives $\mathrm{Coredim}_\Psi(\mathcal{H}_n) \leq n - 1$.

(b), sufficiency. Given a chain with $\mathrm{Domdim}_\Psi(\mathcal{H}_n) < \infty$ for all n, run level selection with

$$d_n = \mathrm{Domdim}_\Psi(\mathcal{H}_n) + 1 \qquad \text{and} \qquad \Phi_s = \Psi(E_s).$$

At a round with $c_s \geq d^\star(h)$ on a stream of $\mathrm{supp}(h)$, the Fact gives $Z_s \subseteq \mathrm{supp}(h)$ and

$$c_s \geq d_{\tilde{n}_s} > \mathrm{Domdim}_\Psi(\mathcal{H}_{\tilde{n}_s}),$$

so the consistent set S_s of c_s distinct points is not a domination witness for $\mathcal{H}_{\tilde{n}_s}$, and Z_s is not finitely dominated. As E_s is finite, $Z_s \not\subseteq \Psi(E_s)$ (Lemma 3.1(b), from anywhere), and the output lies in $Z_s \setminus B_s \subseteq \mathrm{supp}(h) \setminus B_s$. Outputs produced at shallower levels may lie outside Z_s, which is why this direction needs escape from anywhere.

(b), locally Noetherian spans. Then $\mathrm{Coredim}_\Psi = \mathrm{Domdim}_\Psi$ for every class (Remark C.1), and necessity and sufficiency combine. For countable $\mathcal{H}$ the argument of the proof of Proposition 5.1 applies verbatim: the prefix chain if every finite subclass has finite core dimension, and conversely the chain given by necessity, with the core dimension non-increasing under passing to subclasses.

Matroids. For a finitary matroid a set has a finite core iff it has finite rank, and the span is locally Noetherian (Lemma 3.1(d)). So by (a) a countable class is generatable in the limit iff every language has infinite rank, and by (b) a class is non-uniformly generatable iff it is the union of a chain of classes of finite RCdim_M (Corollary 4.2(ii)). The instance for $\Psi = \bigcup_{i \leq k} \mathrm{cl}_{M_i}$ is part of Proposition 7.1. □

E MATROIDS: COORDINATE FLATS, THE PARAPHRASE TRAP, THE HIERARCHY, AND COINCIDENCE

E.1 COORDINATE FLATS

Example E.1 (Coordinate flats)**.** Let $\mathcal{X}$ be the nonzero finitely supported rational sequences with the linear matroid, $c(v)$ the set of nonzero coordinates of v, and $F_J = \{v \in \mathcal{X} : c(v) \subseteq J\}$ for infinite $J \subseteq \mathbb{N}_{\geq 1}$: the vectors that use only coordinates in J, a flat of infinite rank. For a family $\mathcal{J}$ of infinite index sets let $\mathcal{H}_\mathcal{J} = \{\mathbf{1}_{F_J} : J \in \mathcal{J}\}$. A nonempty consistent finite S has safe zone $F_{J^*(S)}$, $J^*(S) = \bigcap\{J \in \mathcal{J} : c(S) \subseteq J\}$, an infinite set of rank $|J^*(S)|$. So $\mathrm{Cdim}(\mathcal{H}_\mathcal{J}) = 0$, while $\mathrm{RCdim}_M(\mathcal{H}_\mathcal{J})$ is 0 if $J^*(\{v\})$ is infinite for every consistent vector v and ∞ otherwise: $\mathcal{H}_\mathcal{J}$ is rank-novel generatable from the first example or not uniformly at all, and in the limit whenever it is countable. If instead $r_M(\mathcal{X}) \leq d$, as for strings placed in $\mathbb{Q}^d$ with the induced linear matroid, at most d rounds can ever succeed: the theory lives on universes of infinite rank.

Proof of the claims in Example E.1. The matroid. Deleting the loop 0 from the linear matroid leaves a finitary matroid on $\mathcal{X}$ with closure $\mathrm{cl}(A) = \mathrm{span}_{\mathbb{Q}}(A) \setminus \{0\}$ (the axioms pass to the restriction; for (CL4), $y \in \mathrm{cl}(A \cup \{x\}) \setminus \mathrm{cl}(A)$ gives $x \in \mathrm{span}(A \cup \{y\})$ with $x \neq 0$).

The flats F_J. If $v \in \mathrm{span}_{\mathbb{Q}}(F_J)$ then every nonzero coordinate of v is a coordinate of some vector of F_J, so $c(v) \subseteq J$: F_J is closed, and the independent vectors e_j ($j \in J$) give $r(F_J) = \infty$. Also $\bigcap_i F_{J_i} = F_{\bigcap_i J_i}$, since $c(v) \subseteq J_i$ for all i iff $c(v) \subseteq \bigcap_i J_i$.

Safe zones. A hypothesis $\mathbf{1}_{F_J}$ is consistent with S iff $c(S) \subseteq J$, so

$$\langle S \rangle_{\mathcal{H}_{\mathcal{J}}} = F_{J^*(S)} \qquad \text{with} \qquad J^*(S) \supseteq c(S) \neq \emptyset;$$

this set contains qv for every $v \in S$ and $q \in \mathbb{Q} \setminus \{0\}$, hence is infinite (here $S \neq \emptyset$ is used), and its rank is $|J^*(S)|$ (basis $\{e_j : j \in J^*(S)\}$).

The two dimensions. Every closure being infinite, $\mathrm{Cdim}(\mathcal{H}_{\mathcal{J}}) = 0$, and by Corollary 4.2(ii),

$$\mathrm{RCdim}_M(\mathcal{H}_{\mathcal{J}}) \in \{\mathrm{Cdim}(\mathcal{H}_{\mathcal{J}}), \infty\} = \{0, \infty\}.$$

Since $J^*(S) \supseteq J^*(\{v\})$ for $v \in S$ (fewer index sets contain $c(S)$ than $c(v)$), a finite realized J^* already occurs at a singleton, so $\mathrm{RCdim}_M \geq 1$ iff some consistent singleton has $|J^*(\{v\})| < \infty$, and $\mathrm{RCdim}_M = 0$ iff $J^*(\{v\})$ is infinite for every consistent v.

Generation. The remaining assertions are Corollary 4.2(ii), Theorem 5.2(a), as every F_J has infinite rank, and the computation $J^*(S) = J_1 \cap J_2$ for $S \subseteq F_{J_1 \cap J_2}$ in Theorem 6.1 (whose class becomes $\mathcal{H}_{\{J_1, J_2\}}$ once the zero vector is deleted; nothing changes since $J_1 \cap J_2 \neq \emptyset$).

Universes of finite rank. Finally, if $r_M(\mathcal{X}) \leq d$ then every independent set has at most d elements (Lemma B.2), and the outputs at successful rounds form an independent set (Remark C.2), so at most d rounds can be successful and no nonempty class is generatable in the limit. □

E.2 The paraphrase trap

Example 4.4 (The paraphrase trap)**.** Partition $\mathcal{X}$ into infinite topics $T_0, T_1, T_2, \ldots$, the paraphrases of one statement each, with the topic matroid, and let $L_1 = T_0 \cup \bigcup_{i \text{ odd}} T_i$ and $L_2 = T_0 \cup \bigcup_{i \geq 2 \text{ even}} T_i$: two languages that share exactly one statement. For $\mathcal{H} = \{\mathbf{1}_{L_1}, \mathbf{1}_{L_2}\}$ every safe zone is T_0, L_1 or L_2, so $\mathrm{Cdim}(\mathcal{H}) = 0$, and paraphrasing T_0 is valid and unseen at every round. Under rank novelty, once an example from T_0 has been seen it spans T_0, so every later novel output lies in only one language, and exhaustion with the core $\{x_1\}$ defeats every pair of thresholds: $\mathcal{H}$ is not non-uniformly generatable. Both languages have infinite rank, so $\mathcal{H}$ is generatable in the limit (Theorem 5.2, Figure 3). The same holds for any two languages of infinite rank, in any finitary matroid, whose intersection is an infinite set of finite rank (Appendix E).

Proof. Setting. We prove the general form. Let M be a finitary matroid and $L_1 \neq L_2$ languages whose intersection $A = L_1 \cap L_2$ is infinite with $r_M(A) < \infty$, and $\mathcal{H} = \{\mathbf{1}_{L_1}, \mathbf{1}_{L_2}\}$.

Safe zones. For finite S: if $S \subseteq A$, then $\mathcal{H}(S) = \mathcal{H}$ and $\langle S \rangle_{\mathcal{H}} = A$; if $S \subseteq L_i$ but $S \not\subseteq A$, then S meets $L_i \setminus L_j$ ($j \neq i$), so $\mathcal{H}(S) = \{\mathbf{1}_{L_i}\}$ and $\langle S \rangle_{\mathcal{H}} = L_i$; otherwise $\mathcal{H}(S) = \emptyset$.

Plain novelty. Every safe zone is therefore A, L_1 or L_2, all infinite, so $\mathrm{Cdim}(\mathcal{H}) = 0$, and

$$G(x_{1:t}) = \mathrm{mindex}(A \setminus \{x_{1:t}\})$$

is valid and unseen at every round.

Rank novelty, fixed thresholds. For rank novelty, given thresholds $d^\star(\cdot)$ let d be the larger of the two, pick a set $S \subseteq A$ of d distinct points and a basis C of A, a finite core by Lemma 3.1(d). As $\mathcal{H}(S) = \mathcal{H}$ and $\langle S \rangle_{\mathcal{H}} = A$, Lemma 3.1(c) defeats the thresholds, so $\mathcal{H}$ is not non-uniformly, hence not uniformly, rank-novel generatable.

Rank novelty in the limit. If both languages have infinite rank, Theorem 5.2(a) makes $\mathcal{H}$ rank-novel generatable in the limit.

The example. In the example, the topic matroid is a finitary matroid (Example 2.2), $L_1 \cap L_2 = T_0$ because the odd indices and the even indices ≥ 2 are disjoint, $r(T_0) = 1$, a single string being a basis of T_0, and each L_i meets infinitely many topics, so it has infinite rank. □

E.3 THEOREM 6.1: STRICT HIERARCHY

Theorem 6.1 (Strict hierarchy). *Within rank-novel generation, uniform ⊆ non-uniform ⊆ limit, and both inclusions are strict for suitable matroids. The first gap is witnessed at the free matroid; the second by the* two-language *class of Example 4.4, whose languages intersect in a single infinite topic, and by two infinite-rank flats of a linear matroid that intersect in a plane. At the free matroid no finite class witnesses the second gap, since finite classes are uniformly generatable there.*

Proof. Inclusions. Uniform ⇒ non-uniform by constant thresholds. Non-uniform ⇒ limit: on a complete enumeration of the infinite $\mathrm{supp}(h)$ the distinct count tends to ∞, so it reaches $d^\star(h)$ at some round, from which the guarantee holds.

First gap (free matroid; the gap itself is classical, Raman et al., 2025). Let $\mathcal{X} = \mathbb{N}$, let O be the odd numbers partitioned into consecutive blocks $W_1, W_2, \ldots$ with $|W_d| = d$, let V be the evens, and

$$\mathcal{H}_1 = \{h_O\} \cup \{a_d : d \geq 1\} \qquad \text{with} \qquad \mathrm{supp}(h_O) = O \quad \text{and} \quad \mathrm{supp}(a_d) = W_d \cup V\text{:}$$

countable, with infinite languages.

First gap, not uniform. The set W_d is consistent exactly with $\{h_O, a_d\}$ (for $j \neq d$, $W_d \subseteq W_j \cup V$ would force $W_d \subseteq W_j$), so $\langle W_d \rangle_{\mathcal{H}_1} = O \cap (W_d \cup V) = W_d$, a finite set of size d: a witness at every d, so $\mathrm{Cdim}(\mathcal{H}_1) = \infty$, and Corollary 4.2(i) applies, the free matroid being the identity span.

First gap, non-uniform. The prefix classes $\mathcal{H}_1^n = \{h_O, a_1, \ldots, a_{n-1}\}$ have as nonempty finite support-intersections exactly the W_j, $j \leq n-1$ (pairs $\{a_i, a_j\}$ intersect in the infinite V), so any witness, sitting inside its own closure, has size at most $n-1$: $\mathrm{Cdim}(\mathcal{H}_1^n) < \infty$, and the sufficiency half of Theorem 5.2(b) applies, since at the identity $\mathrm{Domdim} = \mathrm{Cdim}$ (Remark C.1).

Second gap (structured matroid, finite class). The two-language class of Example 4.4 is one witness, as proved above. For a linear one, let $\mathcal{X} = \mathbb{Q}^{(\omega)}$ be the finitely supported rational sequences with the linear matroid,

$$J_1 = \{1, 2\} \cup \{\text{odd } j \geq 3\}, \qquad J_2 = \{1, 2\} \cup \{\text{even } j \geq 4\},$$

$F_i = \mathrm{span}_{\mathbb{Q}}\{e_j : j \in J_i\}$, and $F_0 = \mathrm{span}_{\mathbb{Q}}(e_1, e_2)$. Coordinatewise, $F_1 \cap F_2 = F_0$, an infinite set of rank 2, while each F_i has infinite rank. Take $\mathcal{H}_2 = \{\mathbf{1}_{F_1}, \mathbf{1}_{F_2}\}$.

Second gap, limit-generatable. By Theorem 5.2(a), as both flats have infinite rank. Directly: on a complete enumeration an example outside F_0 eventually appears; it lies in only one flat, which is then the only consistent hypothesis, of infinite rank, and outputting its least point outside the history span succeeds from then on (Lemma 3.1(b)).

Second gap, not non-uniformly generatable. Given any candidate $(G, d^\star(\cdot))$, let

$$d = \max(d^\star(h_1), d^\star(h_2)) \qquad \text{and} \qquad S = \{e_1, 2e_1, \ldots, de_1\} \subseteq F_0.$$

Both hypotheses are consistent with S, so $\mathcal{H}_2(S) = \mathcal{H}_2$ and

$$A = \langle S \rangle_{\mathcal{H}_2} = F_1 \cap F_2 = F_0,$$

which has the finite core $C = \{e_1, e_2\} \subseteq F_0$. Since $d^\star(h) \leq |S|$ for both $h \in \mathcal{H}_2(S)$, Lemma 3.1(c) (non-uniform form) defeats $(G, d^\star(\cdot))$: the output after S, e_1, e_2 is either in $F_0 \subseteq \mathrm{span}(C) \subseteq B_n$ or outside $F_0 = F_1 \cap F_2$, hence outside some F_i, whose hypothesis the adversary commits to.

No finite witness at the free matroid. A finite class witnessing this gap is impossible at the free matroid, where finite classes have finite Cdim (proof of Proposition 5.1) and are uniformly generatable (Corollary 4.2(i)). □

E.4 THEOREM 6.2: WHEN THE THEORIES COINCIDE

Theorem 6.2 (When the theories coincide). *For a finitary matroid M on $\mathcal{X}$ the following are equivalent:* (a) $\mathrm{RCdim}_M(\mathcal{H}) = \mathrm{Cdim}(\mathcal{H})$ *for every class $\mathcal{H}$ on $\mathcal{X}$;* (b) *every finite-rank flat of M is finite;* (c) *M has no infinite set of finite rank. Under these conditions every level of rank-novel generation coincides, as a property of classes, with the classical one (at the limit level, for countable*

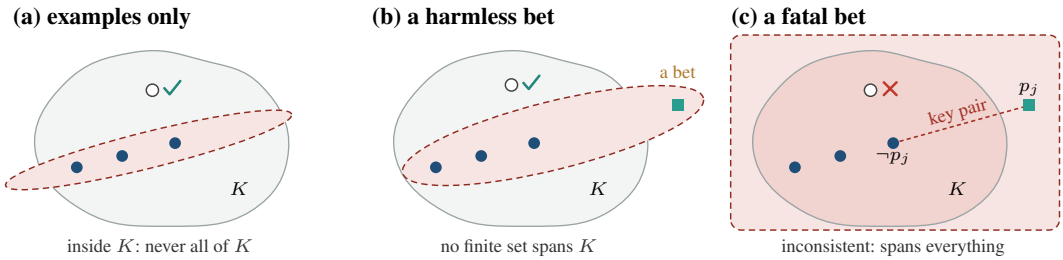

Figure 5: Inside and outside, drawn as in Figure 1: dots are examples, squares the generator's own outputs, and the red region is the span of the history. (a) Finitely many points of K never span a K without a finite core. (b) If the span is locally Noetherian, no finite set spans such a K, not even one with a bet outside K, so the bet does no harm. (c) With explosion, the bet p_j and the example $\neg p_j$ form a key pair: they are inconsistent, and their span is everything.

classes), and every M violating them carries a one-language class realizing the separation of Proposition 4.3(ii).

Proof. (b)⇔(c). $\mathrm{cl}_M(A)$ is a flat by (CL3). If every finite-rank set is finite, so is every finite-rank flat. Conversely, if $r_M(A) = k < \infty$ then $A \subseteq \mathrm{cl}_M(F_0)$ with $|F_0| = k$, a flat of rank at most k (Lemma B.3(F)), finite by (b).

(c)⇒(a). Under (c), a set is finite iff it has finite rank, so the Cdim and RCdim_M witness conditions agree pointwise for every class.

¬(c)⇒ ¬(a). Let A_0 be infinite with $r_M(A_0) < \infty$ and $\mathcal{H} = \{\mathbf{1}_{A_0}\}$. Every consistent set has closure A_0: infinite, so $\mathrm{Cdim}(\mathcal{H}) = 0$; of finite rank and with arbitrarily many distinct points, so $\mathrm{RCdim}_M(\mathcal{H}) = \infty$. The argument of Proposition 4.3(ii) runs verbatim with $\mathbb{Q}^2$ replaced by A_0: plainly uniformly generatable from one example, not rank-novel generatable even in the limit (Theorem 5.2(a); a basis of A_0 is a finite core of the language, Lemma 3.1(d)).

Levels coincide under (a)–(c). By Lemma 3.1(d), the core dimension of the matroid span cl_M is RCdim_M and that of the identity is Cdim, so by (a) the two spans have the same core dimension for every class; both spans are locally Noetherian (Lemma 3.1(d)). Plain novelty is the identity with the examples-only span, so Proposition 5.1 applies to it at every level.

Uniform, and success at every round. By Theorem 4.1 for cl_M and by Proposition 5.1 for plain novelty, uniform generatability holds iff the core dimension is finite, with optimal threshold one more, and in both cases success at every round is threshold 1, possible iff the core dimension is 0 (proof of Theorem 4.1). So rank-novel and plain generation agree, with the same optimal threshold $\mathrm{Cdim}(\mathcal{H}) + 1$.

Non-uniform. By Theorem 5.2(b) for the locally Noetherian span cl_M, and by Proposition 5.1 for plain novelty, each is characterized by the existence of a chain of classes of finite core dimension, and by (a), applied to every class of the chain, the two conditions agree.

Limit, countable classes. Under (c) every language, being infinite, has infinite rank, so by Theorem 5.2(a) every countable class is rank-novel generatable in the limit, and so it is plainly (Proposition 4.3(i), or Proposition 5.1 at the identity). The free matroid satisfies (c), and for it these coincidences are the classical theory. We make no claim for uncountable classes in the limit game; for plain novelty they are characterized by finite witnesses (Li et al., 2026b). □

F Locally Noetherian spans

The name follows the locally Noetherian categories of Gabriel (1962), generated by Noetherian objects; here the finitely generated closed sets play that role. The global condition would be far too strong: the identity on infinitely many strings is locally Noetherian, although its spans form infinite increasing chains.

F.1 THEOREM 5.3: THE RANK FORM

Theorem 5.3 (The rank form needs exactly local Noetherianity)**.** *For a finitary closure space $(\mathcal{X}, \mathrm{cl})$ the following are equivalent:* (i) *it is locally Noetherian;* (ii) *for every class $\mathcal{H}$, $\mathcal{H}$ is uniformly generatable iff $\mathrm{Domdim}_{\mathrm{cl}}(\mathcal{H}) < \infty$, and then the optimal threshold is $\mathrm{Domdim}_{\mathrm{cl}}(\mathcal{H}) + 1$;* (iii) *for every countable class $\mathcal{H}$, $\mathcal{H}$ is generatable in the limit iff no language of $\mathcal{H}$ is finitely dominated. The sufficiency halves of (ii) and (iii) hold in every finitary closure space, and each necessity half, even restricted to one-language classes, already implies (i).*

Proof. The sufficiency halves hold in every finitary closure space. For (ii), let $D = \mathrm{Domdim}_{\mathrm{cl}}(\mathcal{H}) < \infty$ and output $\mathrm{mindex}(\langle x_{1:s}\rangle_{\mathcal{H}} \setminus \mathrm{cl}(E_s))$ (totality convention). At a round with $c_s \geq D + 1$ on a stream of $\mathrm{supp}(h)$, the set S_s is consistent with $c_s > D$ distinct points, so its safe zone is not finitely dominated, and Lemma 3.1(b), from anywhere, gives a candidate outside $\mathrm{cl}(E_s) = B_s$; the output lies in the safe zone, inside $\mathrm{supp}(h)$. So threshold $D + 1$ is achieved. For (iii), this is the escape side of Theorem 5.2(a). Neither argument uses a closure axiom.

(i)⇒(ii). Under local Noetherianity $\mathrm{Coredim}_{\mathrm{cl}} = \mathrm{Domdim}_{\mathrm{cl}}$ for every class (Remark C.1), and Theorem 4.1 gives (ii).

(i)⇒(iii). This is the last sentence of Theorem 5.2(a), since under local Noetherianity a language is finitely dominated iff it has a finite core.

¬(i)⇒ failure of both necessity halves. Suppose $A \subseteq \mathrm{cl}(F)$ for a finite F but A has no finite core. Then A is infinite, since a finite A satisfies $A \subseteq \mathrm{cl}(A)$ by (CL1).

The class $\mathcal{H}_A$. Let $\mathcal{H}_A = \{\mathbf{1}_A\}$, a one-language class. For every d, any d distinct elements of A form a consistent set whose safe zone is A (the class is a singleton), finitely dominated by F: so $\mathrm{Domdim}_{\mathrm{cl}}(\mathcal{H}_A) = \infty$, and the language A is finitely dominated.

The generator G_A. Define

$$G_A(x_{1:t}) = \mathrm{mindex}(A \setminus \mathrm{cl}(E_t))$$

when the entries lie in A and the set is nonempty, and ξ_1 otherwise. On a legal stream all $x_i \in A$; by induction on t, if $\hat{x}_{<t} \in A$ then E_t is a finite subset of A, so $A \not\subseteq \mathrm{cl}(E_t) = B_t$ because A has no finite core (Lemma 3.1(b), from inside); the candidate set is nonempty, and $\hat{x}_t \in A \setminus B_t$ is valid, novel, and in A, which closes the induction.

Conclusion. Thus G_A succeeds at every round $t \geq 1$ on every legal stream: uniform generation at threshold 1 (as $c_1 = 1$) and generation in the limit with $t^* = 1$. So the necessity half of (ii) fails for $\mathcal{H}_A$, and so does the necessity half of (iii). Contrapositively, each necessity half, even for one-language classes, implies (i). □

F.2 PROPOSITION 5.4: LOCALLY NOETHERIAN SPANS

Proposition 5.4 (Locally Noetherian spans)**.** (i) *A span is locally Noetherian iff no finitely dominated set contains an infinite sequence $o_1, o_2, \ldots$ with $o_i \notin \Psi(\{o_1, \ldots, o_{i-1}\})$ for all i.* (ii) *For a finitary closure operator, a set has a finite core iff its closure is finitely generated, and the span is locally Noetherian iff every increasing chain of closed sets inside any finitely generated closed set is eventually constant: the ascending chain condition (Noether, 1921) below every finitely generated closed set. Exchange and bounded such sequences are strictly stronger conditions (Proposition F.1).*

Proof. An *escape sequence in A* is a finite or infinite sequence $o_1, o_2, \ldots$ of elements of A with $o_i \notin \Psi(\{o_1, \ldots, o_{i-1}\})$ for every i (for $i = 1$: $o_1 \notin \Psi(\emptyset)$).

(i), only if. Let Ψ be locally Noetherian, and suppose $A \subseteq \Psi(F)$, F finite, contains an infinite escape sequence $O = \{o_1, o_2, \ldots\}$. As $O \subseteq A$ is finitely dominated, it has a finite core $C \subseteq O$, and $C \subseteq \{o_1, \ldots, o_m\}$ for some m; by monotonicity

$$o_{m+1} \in O \subseteq \Psi(C) \subseteq \Psi(\{o_1, \ldots, o_m\}),$$

which contradicts the escape property.

(i), if. Conversely, suppose that no finitely dominated set contains an infinite escape sequence, and let A be finitely dominated. If $A \subseteq \Psi(\emptyset)$, then $\emptyset$ is a finite core. Otherwise let $o_1 = \mathrm{mindex}(A \setminus \Psi(\emptyset))$ and, while $A \not\subseteq \Psi(\{o_1, \dots, o_j\})$, let

$$o_{j+1} = \mathrm{mindex}(A \setminus \Psi(\{o_1, \dots, o_j\})).$$

This is an escape sequence in A, hence finite: the recursion stops at some j with $A \subseteq \Psi(\{o_1, \dots, o_j\})$, and $\{o_1, \dots, o_j\}$ is a finite core of A.

(ii) Let $\Psi = \mathrm{cl}$ be a finitary closure operator. Call D a *flat* (closed) if $\mathrm{cl}(D) = D$, and *finitely generated* if $D = \mathrm{cl}(F)$ for a finite F. The union D of an increasing sequence of flats $D_1 \subseteq D_2 \subseteq \cdots$ is a flat: by (CL5), $\mathrm{cl}(D) = \bigcup\{\mathrm{cl}(F) : F \subseteq D \text{ finite}\}$, each such F lies in some D_j, so $\mathrm{cl}(F) \subseteq \mathrm{cl}(D_j) = D_j \subseteq D$ by (CL2), and $D \subseteq \mathrm{cl}(D)$ by (CL1).

Cores. If $C \subseteq A$ is a finite core, then

$$\mathrm{cl}(C) \subseteq \mathrm{cl}(A) \subseteq \mathrm{cl}(\mathrm{cl}(C)) = \mathrm{cl}(C)$$

by (CL2) and (CL3), so $\mathrm{cl}(A) = \mathrm{cl}(C)$ is finitely generated. Conversely, let $\mathrm{cl}(A) = \mathrm{cl}(F)$ with F finite. By (CL5) each $f \in F \subseteq \mathrm{cl}(A)$ lies in $\mathrm{cl}(F_f)$ for a finite $F_f \subseteq A$; then $C = \bigcup_{f \in F} F_f \subseteq A$ is finite, $F \subseteq \mathrm{cl}(C)$, and

$$A \subseteq \mathrm{cl}(A) = \mathrm{cl}(F) \subseteq \mathrm{cl}(\mathrm{cl}(C)) = \mathrm{cl}(C):$$

C is a finite core. Likewise $A \subseteq \mathrm{cl}(F)$ iff $\mathrm{cl}(A) \subseteq \mathrm{cl}(F)$, by (CL1)–(CL3), so A is finitely dominated iff $\mathrm{cl}(A)$ lies in a finitely generated flat.

The chain condition. Suppose cl is locally Noetherian, and let $D_1 \subseteq D_2 \subseteq \cdots$ be flats inside a finitely generated flat c. Their union D is a flat inside c, hence finitely dominated, hence with a finite core, hence finitely generated, $D = \mathrm{cl}(F')$ with F' finite; F' lies in some D_j, so

$$D = \mathrm{cl}(F') \subseteq D_j \subseteq D,$$

and the sequence is constant from j on. Conversely, assume the chain condition, and let A be finitely dominated, so that $D = \mathrm{cl}(A)$ is a flat inside a finitely generated flat c. If D were not finitely generated, the recursion

$$F_0 = \emptyset, \qquad F_{i+1} = F_i \cup \{\mathrm{mindex}(D \setminus \mathrm{cl}(F_i))\}$$

would never stop, since $\mathrm{cl}(F_i) \subseteq D$ and $\mathrm{cl}(F_i) \neq D$ for all i, and

$$\mathrm{cl}(F_0) \subsetneq \mathrm{cl}(F_1) \subsetneq \cdots$$

would be a strictly increasing sequence of flats inside c. So $\mathrm{cl}(A)$ is finitely generated, and A has a finite core by the previous paragraph. □

F.3 The hierarchy of spans

Proposition F.1 (Matroids and the hierarchy)**.** (i) *In a finitary matroid, escape sequences inside A have length at most $r_M(A)$, so every matroid is locally Noetherian.* (ii) *Among finitary closure spaces, exchange, bounded escape chains and local Noetherianity are strictly weaker in turn.*

Proof. (i) In a finitary matroid an escape sequence in A is independent (Lemma B.1), so it has at most $r_M(A)$ elements (Lemma B.3(C)). A finitely dominated set has finite rank, so by Proposition 5.4(i) every finitary matroid is locally Noetherian.

(ii) Say that a finitary closure space has *bounded escape chains* (BEC) if every finitely dominated A bounds the lengths of the escape sequences in A. By (i) every finitary matroid has BEC, with bound $r_M(A)$, and BEC implies local Noetherianity by Proposition 5.4(i), since the initial segments of an infinite escape sequence are escape sequences of every length.

The witnesses. The following spaces (Figure 6) show that neither implication reverses and that not every finitary closure space is locally Noetherian. Let $\mathcal{X}_0 = \{a, b, z_1, z_2, \dots\}$, written $\mathcal{X}_1$ when it carries cl_1, and call a set *hot* if it contains both a and b, and *cold* otherwise.

The witness $(\mathcal{X}_1, \mathrm{cl}_1)$: BEC without exchange. Here

$$\mathrm{cl}_1(A) = A \cup \{z_1\} \text{ for hot } A \qquad \text{and} \qquad \mathrm{cl}_1(A) = A \text{ otherwise.}$$

(CL1) is clear; (CL2): a hot $A \subseteq A'$ forces a hot A' and $A \cup \{z_1\} \subseteq A' \cup \{z_1\}$, while a cold A is contained in either image of A'; (CL3): cold sets are fixed, and for hot A the set $A \cup \{z_1\}$ is hot and adding z_1 again changes nothing; (CL5): for hot A the finite hot subset $\{a, b\}$ contributes $\{a, b, z_1\}$ and singletons contribute the rest, while for cold A every finite subset is cold and their images union to A.

Exchange fails. $z_1 \in \mathrm{cl}_1(\{a, b\}) \setminus \mathrm{cl}_1(\{a\})$ but $b \notin \mathrm{cl}_1(\{a, z_1\}) = \{a, z_1\}$.

BEC *holds.* $\mathrm{cl}_1(F) \subseteq F \cup \{z_1\}$ is finite for finite F, so a set is finitely dominated iff it is finite, and an escape sequence in a finite A has pairwise distinct elements (by (CL1)), hence length at most $|A|$.

The witness $(\mathbb{N}_0, \mathrm{cl}_\uparrow)$*: locally Noetherian without* BEC. Here

$$\mathrm{cl}_\uparrow(\emptyset) = \emptyset \qquad \text{and} \qquad \mathrm{cl}_\uparrow(A) = \{n \geq \min A\} \text{ otherwise.}$$

(CL1): $A \subseteq \{n \geq \min A\}$. (CL2): if $\emptyset \neq A \subseteq A'$ then $\min A' \leq \min A$, so $\mathrm{cl}_\uparrow(A) \subseteq \mathrm{cl}_\uparrow(A')$; the case $A = \emptyset$ is trivial. (CL3): $\min \mathrm{cl}_\uparrow(A) = \min A$. (CL5): $\mathrm{cl}_\uparrow(A) = \mathrm{cl}_\uparrow(\{\min A\})$ for $A \neq \emptyset$.

Locally Noetherian. Every nonempty A satisfies $A \subseteq \mathrm{cl}_\uparrow(\{\min A\})$ with $\{\min A\} \subseteq A$.

Exchange fails. $1 \in \mathrm{cl}_\uparrow(\{0\}) \setminus \mathrm{cl}_\uparrow(\emptyset)$ but $0 \notin \mathrm{cl}_\uparrow(\{1\}) = \{n \geq 1\}$.

BEC *fails.* It fails on $\mathbb{N}_0$, which $\{0\}$ dominates: for each N the sequence $N, N-1, \ldots, 0$ is an escape sequence, since each entry is smaller than the minimum of its predecessors and hence outside their upward closure, and the first entry lies outside $\mathrm{cl}_\uparrow(\emptyset) = \emptyset$; the lengths $N+1$ are unbounded.

In the form (ii). The flats are $\emptyset$ and the sets $[m, \infty)$, an increasing sequence of them is eventually constant because the minima form a non-increasing sequence of naturals, and the chains

$$[N, \infty) \subsetneq [N-1, \infty) \subsetneq \cdots \subsetneq [0, \infty) = \mathrm{cl}_\uparrow(\{0\})$$

have unbounded length.

Generation. The class $\{\mathbf{1}_{\mathbb{N}_0}\}$ is generatable at no level, although BEC fails on the very set that the game realizes: $\{0\}$ is a finite core of $\mathbb{N}_0$, and every set of distinct points is a consistent set with safe zone $\mathbb{N}_0$, so Lemma 3.1(c) applies at every threshold and in the limit; directly, once 0 has been played, $B_t \supseteq \mathrm{cl}_\uparrow(\{0\}) = \mathbb{N}_0$.

The witness $(\mathcal{X}_0, \mathrm{cl}_0)$*: not locally Noetherian.* Here

$$\mathrm{cl}_0(A) = \mathcal{X}_0 \text{ for hot } A \qquad \text{and} \qquad \mathrm{cl}_0(A) = A \text{ otherwise.}$$

(CL1) is clear; (CL2): a hot $A \subseteq A'$ forces a hot A', and both map to $\mathcal{X}_0$, while a cold A satisfies $\mathrm{cl}_0(A) = A \subseteq A' \subseteq \mathrm{cl}_0(A')$; (CL3): $\mathcal{X}_0$ and cold sets are fixed; (CL5): for hot A the finite subset $\{a, b\}$ already maps to $\mathcal{X}_0$, while for cold A every finite subset is cold and their images union to A.

Exchange fails. $z_1 \in \mathrm{cl}_0(\{a, b\}) \setminus \mathrm{cl}_0(\{a\})$ but $b \notin \mathrm{cl}_0(\{a, z_1\}) = \{a, z_1\}$.

A finitely dominated set without a finite core. The set $K_0 = \{z_1, z_2, \ldots\} \subseteq \mathrm{cl}_0(\{a, b\}) = \mathcal{X}_0$ is finitely dominated, but every finite $C \subseteq K_0$ is cold (as $a, b \notin K_0$), so $\mathrm{cl}_0(C) = C \not\supseteq K_0$: K_0 has no finite core.

In the form (ii). Cold sets are flats, and

$$\{z_1\} \subsetneq \{z_1, z_2\} \subsetneq \cdots$$

never stops inside the finitely generated flat $\mathcal{X}_0 = \mathrm{cl}_0(\{a, b\})$.

Generation. The class $\{\mathbf{1}_{K_0}\}$ is the ¬(i) direction of Theorem 5.3 at $A = K_0$: the generator "least-indexed z_j outside the history" never lets the history become hot, so $B_s = E_s$ and a fresh z always escapes.

The lattices below finitely generated flats. The flats of a finitary closure space form a complete lattice under inclusion, with intersections as meets and $\mathrm{cl}(\bigcup_i D_i)$ as joins; it is algebraic, and its compact elements are the finitely generated flats (Davey & Priestley, 2002, Ch. 7). The flats inside a finitely generated flat c form the lattice $\downarrow c$, and by (ii) local Noetherianity is the ascending chain condition in every such lattice.

Escape sequences and chains. An escape sequence $o_1, \ldots, o_m$ in a flat c gives the strictly increasing chain of flats

$$\mathrm{cl}(\emptyset) \subsetneq \mathrm{cl}(\{o_1\}) \subsetneq \cdots \subsetneq \mathrm{cl}(\{o_1, \ldots, o_m\})$$

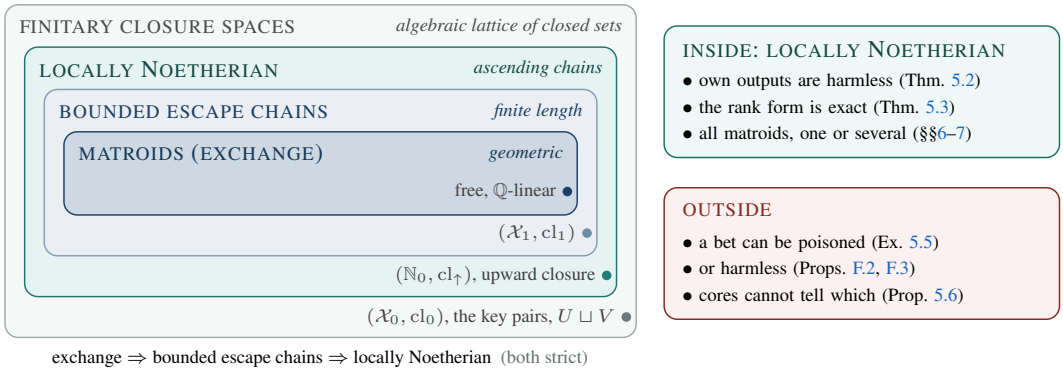

Figure 6: Locally Noetherian spans among finitary closure spaces (Propositions 5.4 and F.1). Each ring names a condition on the space (left) and what it says about the lattice of closed sets below each finitely generated one (right, in italics; the whole lattice is algebraic). Bounded escape chains (BEC) and local Noetherianity (FID) are exactly finite length and ascending chains; exchange makes these lattices geometric but is not a lattice condition. The pin in each band is a space that satisfies the ring's condition but not the next one's (proof of Proposition F.1). The cards summarize what happens inside and outside local Noetherianity; uniform generation is characterized in every space by the core dimension (Theorem 4.1).

inside c, and conversely a strictly increasing chain

$$D_0 \subsetneq D_1 \subsetneq \cdots \subsetneq D_m$$

of flats inside c gives the escape sequence $o_1, \ldots, o_m$ in c with $o_i \in D_i \setminus D_{i-1}$, since $\mathrm{cl}(\{o_1, \ldots, o_{i-1}\}) \subseteq D_{i-1}$ and $\mathrm{cl}(\emptyset) \subseteq D_0$. Every finitely dominated set lies in a finitely generated flat, and finitely generated flats are finitely dominated, so BEC says exactly that every $\downarrow c$ has finite length.

Matroids give geometric lattices. In a matroid, $\downarrow c$ has length at most $r_M(c)$ by (i), and exchange makes it a *geometric* lattice, one of finite length that is also semimodular and atomistic (Oxley, 2011). Indeed, for a flat $D \subseteq c$ and $x \in c \setminus D$ the flat $\mathrm{cl}_M(D \cup \{x\})$ covers D: a flat E with $D \subsetneq E \subseteq \mathrm{cl}_M(D \cup \{x\})$ contains some $y \notin D$, and exchange gives $x \in \mathrm{cl}_M(D \cup \{y\}) \subseteq E$. Hence the covers of D are exactly these flats, the atoms are the $\mathrm{cl}_M(\{x\})$ with $x \notin \mathrm{cl}_M(\emptyset)$, and every flat $D \subseteq c$ is the join of the atoms $\mathrm{cl}_M(\{x\})$, x in a basis of D (Lemma 3.1(d)). If D covers $D \cap D'$, then $D = \mathrm{cl}_M((D \cap D') \cup \{x\})$ for any $x \in D \setminus D'$, so the join $\mathrm{cl}_M(D \cup D') = \mathrm{cl}_M(D' \cup \{x\})$ covers D': this is semimodularity.

Exchange is not a lattice condition. Exchange is nevertheless not a condition on these lattices: on $\{u, v, w\}$ the closed sets $\emptyset$, $\{u\}$, $\{v\}$ and $\{u, v, w\}$ define a finite, hence finitary, closure space whose lattice is that of the free matroid on two points, while exchange fails, as $u \in \mathrm{cl}(\{w\}) \setminus \mathrm{cl}(\emptyset)$ but $w \notin \mathrm{cl}(\{u\}) = \{u\}$. What is proved above is Proposition 5.4, part (i), the implications and the witnesses; the lattice vocabulary is a reading of them. □

F.4 EXPLOSION

Example 5.5 (Explosion)**.** Let the strings be the literals $p_j, \neg p_j$ ($j \geq 1$), and let a set of literals span every literal it entails in classical logic: itself if it contains no pair $\{p_j, \neg p_j\}$, and everything otherwise. The languages are the literals true under an assignment that makes only finitely many variables false. No language has a finite core, yet every language lies in the span of one pair: the span is not locally Noetherian. Every finite subclass is generatable from round one (Theorem 4.1), but the whole class is not generatable in the limit. An adversary that simulates the generator on the language of true literals finds its first success, with a literal p_j, and commits in advance to the language with $\neg p_j$, enumerated identically up to that round and then $\neg p_j$; from then on the history is inconsistent, spans everything, and no output is ever novel again. Counting only the examples as said, the class is generatable in the limit (Proposition 5.1).

Proof. The span is consequence on literals. Write $a_j = p_j$ and $b_j = \neg p_j$, so that $\mathcal{X} = \{a_j, b_j : j \geq 1\}$, and read $\neg\neg p_j$ as p_j. A set Γ of literals entails each of its members. If Γ contains no complementary pair and $\ell \notin \Gamma$ is a literal, then $\Gamma \cup \{\neg\ell\}$ contains no complementary pair either, since the complement of $\neg\ell$ is $\ell \notin \Gamma$; a set of literals without a complementary pair is satisfiable (make each of its literals true), so $\Gamma \not\models \ell$. If Γ contains a pair $\{p_j, \neg p_j\}$, it has no model and entails every literal. So the span is

$$\mathrm{cl}(E) = \mathcal{X} \text{ if } E \text{ contains a } \textit{key pair } \{a_j, b_j\}, \quad \text{and} \quad \mathrm{cl}(E) = E \text{ otherwise.}$$

The same computation with the literals $p, \neg p, q_1, q_2, \ldots$ and $a = p$, $b = \neg p$, $z_i = q_i$ gives the space $(\mathcal{X}_0, \mathrm{cl}_0)$ of the proof of Proposition F.1.

The languages. For finitely supported $\sigma \in \{0,1\}^{\mathbb{N}_{\geq 1}}$ let

$$K_\sigma = \{a_j : \sigma_j = 0\} \cup \{b_j : \sigma_j = 1\}:$$

the literals true under the assignment that makes p_j false exactly when $\sigma_j = 1$. Every assignment that makes only finitely many variables false arises from exactly one such σ, so the class of the example is $\mathcal{H} = \{\mathbf{1}_{K_\sigma}\}$, a countable class of infinite languages.

The span is not locally Noetherian. The axioms: (CL1) is clear; if $E \subseteq F$ and E contains a key pair, so does F, and otherwise $\mathrm{cl}(E) = E \subseteq \mathrm{cl}(F)$; $\mathcal{X}$ and the sets without key pairs are fixed; and a set containing a key pair P has $\mathrm{cl}(P) = \mathcal{X}$, while one without is the union of its finite subsets, which are fixed.

No finite cores, yet finitely dominated. Each K_σ contains exactly one point of every key pair, so every finite $C \subseteq K_\sigma$ has $\mathrm{cl}(C) = C \not\supseteq K_\sigma$, although $K_\sigma \subseteq \mathrm{cl}(\{a_1, b_1\}) = \mathcal{X}$: no language has a finite core, every language is finitely dominated, and the span is not locally Noetherian.

Finite subclasses. For a finite subclass, the safe zone of a consistent set is an intersection of some K_σ's; it contains no key pair, so $\mathrm{cl}(C) = C$ for its finite subsets C, and it contains a_j for every j beyond the supports involved, so it is infinite. It has no finite core, the core dimension is 0, and by Theorem 4.1 some generator is valid and novel at every round.

The whole class, with own outputs counted. Let G be any generator and enumerate K_0 ($\sigma \equiv 0$) as $a_1, a_2, \ldots$. If G fails infinitely often, it does not generate K_0 in the limit. Otherwise it succeeds at some round t, with an output $a_j \notin \mathrm{cl}(E_t) \supseteq \{a_1, \ldots, a_t\}$, so $j > t$.

The adversary's stream. Let σ' have a single 1, at j, and enumerate $K_{\sigma'}$ as $a_1, \ldots, a_t, b_j$ followed by the remaining a_i. This stream agrees with the first on its first t entries, so G outputs a_j at round t as before; the example b_j follows, and from round $t+1$ on the history contains the key pair $\{a_j, b_j\}$, spans $\mathcal{X}$, and no round succeeds. So no generator generates $\mathcal{H}$ in the limit.

Counting only the examples. On a stream of a language K_σ the examples lie in K_σ, which contains no key pair, so $\mathrm{cl}(S_s) = S_s$ and success against the examples-only span is plain novelty. $\mathcal{H}$ is countable and no language has a finite core, so Proposition 5.1 makes $\mathcal{H}$ generatable in the limit against this span. □

F.5 Master keys

Beyond locally Noetherian spans the core form of Theorem 5.2 can also survive. In $(\mathcal{X}_0, \mathrm{cl}_0)$ the two points a and b are master keys that together span everything, and a generator that never says a key keeps the core form at every level.

Proposition F.2 (Master keys)**.** *In the space $(\mathcal{X}_0, \mathrm{cl}_0)$ of the proof of Proposition F.1, which is not locally Noetherian, and with the full history $B_s = \mathrm{cl}_0(E_s)$: a set has a finite core iff it is finite or contains both a and b; a countable class is generatable in the limit iff no language has a finite core; and a class is non-uniformly generatable iff it is the union of a chain of classes of finite core dimension.*

Proof. Cores. If $\{a, b\} \subseteq A$, then $\{a, b\}$ is a finite core of A, as $\mathrm{cl}_0(\{a, b\}) = \mathcal{X}_0$; a finite set is its own core; and if A is infinite and misses a or b, every finite $C \subseteq A$ misses it too, so $\mathrm{cl}_0(C) = C \not\supseteq A$. Hence a set without a finite core is infinite and contains at most one of a, b, so it contains infinitely many of the z_i.

Key observation. A set E that contains at most one of a, b has

$$\mathrm{cl}_0(E) = E.$$

So if all outputs of a generator are z's and the examples contain at most one of a, b, then $B_s = E_s$ at every round.

Limit. Necessity is the exhaustion side of Theorem 5.2(a). Conversely, let no language have a finite core, and use the scaffold of Appendix D: output the least-indexed $z_i \in L_{n_t} \setminus E_t$ when $\mathrm{Crit}_t \neq \emptyset$ and such a z_i exists, and z_1 otherwise.

Limit, success. All outputs are z's, and the examples lie in the target K, which has no finite core and so contains at most one of a, b; hence $B_t = E_t$. For $t \geq t^*$ the language $L_{n_t} \subseteq K$ has no finite core, so it contains infinitely many z's and some $z_i \in L_{n_t} \setminus E_t$ exists; the output is valid, as $L_{n_t} \subseteq K$, and novel, as $z_i \notin E_t = B_t$.

Non-uniform. Necessity is Theorem 5.2(b), which holds for every span.

Non-uniform, no language holds both keys. Conversely, let $\mathcal{H} = \bigcup_n \mathcal{H}_n$ for a chain with $\mathrm{Coredim}(\mathcal{H}_n) < \infty$ for all n. No language of $\mathcal{H}$ contains both a and b: if $h \in \mathcal{H}_m$ had, every finite S with $\{a, b\} \subseteq S \subseteq \mathrm{supp}(h)$ would be consistent with $\mathcal{H}_m$ and have $\{a, b\} \subseteq S \subseteq \langle S \rangle_{\mathcal{H}_m}$, a safe zone with a finite core, and such S of every size exist since $\mathrm{supp}(h)$ is infinite; so $\mathrm{Coredim}(\mathcal{H}_m) = \infty$.

Non-uniform, the generator. Run level selection (Appendix D) with $d_n = \mathrm{Coredim}(\mathcal{H}_n) + 1$, outputting the least-indexed $z_i \in Z_s \setminus E_s$, and z_1 whenever this rule does not apply. All outputs are z's, and on every legal stream the examples contain at most one of a, b, so $B_s = E_s$.

Non-uniform, success. At a round with $c_s \geq d^\star(h)$ on a stream of $\mathrm{supp}(h)$, the Fact of level selection gives $Z_s \subseteq \mathrm{supp}(h)$ and

$$c_s \geq d_{\tilde{n}_s} > \mathrm{Coredim}(\mathcal{H}_{\tilde{n}_s}),$$

so Z_s has no finite core, as S_s would otherwise be a core witness for $\mathcal{H}_{\tilde{n}_s}$ of size c_s. So Z_s contains infinitely many z's, some $z_i \in Z_s \setminus E_s$ exists, and the output is valid and novel. □

F.6 Same inside, opposite answers

What the languages cannot see. The two situations beyond local Noetherianity look the same from inside the languages, and one principle covers both positive results.

Proposition 5.6 (Same inside, opposite answers). *Let $\mathcal{X}_0 = \{a, b, z_1, z_2, \ldots\}$ with $\mathrm{cl}_0(E) = \mathcal{X}_0$ if $\{a, b\} \subseteq E$ and $\mathrm{cl}_0(E) = E$ otherwise, and send the literals of Example 5.5 into it by $\varphi(p_j) = z_{2j-1}$, $\varphi(\neg p_j) = z_{2j}$. Then φ carries the languages of Example 5.5 onto those of a class $\mathcal{H}^\varphi$, and every set A inside a single language is its own span, as is $\varphi(A)$, and both are finitely dominated. With the generator's own outputs counted, $\mathcal{H}^\varphi$ is generatable in the limit and non-uniformly, and the class of Example 5.5 at neither level. So beyond locally Noetherian spans no criterion that sees only the languages and the sets inside single languages, up to such a relabelling, decides these levels.*

Precisely: no property Φ of a space, a span and a class that takes the same value on two triples whenever a bijection ψ between the unions U, U' of their languages carries languages onto languages, satisfies

$$\psi(\Psi(A) \cap U) = \Psi'(\psi(A)) \cap U'$$

and preserves finite domination for every A inside a single language, can agree with generatability in the limit (or non-uniform generatability) on both triples of the proposition, with the full history.

Proof. The relabelling preserves the inside. φ is a bijection from the literals onto $\{z_1, z_2, \ldots\}$. Each K_σ contains exactly one literal of every key pair, so a set A inside a language contains no key pair and $\mathrm{cl}(A) = A$; $\varphi(A)$ contains neither key, so $\mathrm{cl}_0(\varphi(A)) = \varphi(A)$; and $\mathrm{cl}(\{a_1, b_1\})$ and $\mathrm{cl}_0(\{a, b\})$ are the whole spaces. Consistent sets, safe zones (which lie inside consistent languages) and sizes correspond under φ, so the listed properties take the same value on both sides, for every subclass.

The key-pair class. The class $\mathcal{H}$ is not generatable in the limit (Example 5.5), hence not non-uniformly either: on a complete enumeration of an infinite language the count of distinct examples passes every threshold, so non-uniform generation implies generation in the limit.

The relabelled class, in the limit. The languages of $\mathcal{H}^{\varphi}$ are infinite and contain neither key, so none has a finite core, and Proposition F.2 gives generation in the limit.

The relabelled class, non-uniformly. A finite subclass has as safe zones intersections of finitely many $\varphi(K_\sigma)$, each containing z_{2j-1} for all but finitely many j; they are infinite, contain neither key, and have no finite core, so the core dimension of every finite subclass is 0, and Proposition F.2 gives non-uniform generation along the chain of initial segments. □

Sets that straddle two languages do tell the two apart: $K_\sigma \cup K_\tau$ has the finite core $\{a_j, b_j\}$ whenever $\sigma_j \neq \tau_j$, while $\varphi(K_\sigma) \cup \varphi(K_\tau)$ has none. A characterization beyond local Noetherianity must look at such sets.

F.7 Safe bets and a region chosen after the first example

Proposition F.3 (Safe bets)**.** *Let Ψ be any map on subsets of $\mathcal{X}$, let $B_s = \Psi(E_s)$, and let $\mathcal{H}$ be countable. Call $P \subseteq \mathcal{X}$* safe *if $L \cap P \not\subseteq \Psi(E)$ whenever $L \subseteq K$ are languages of $\mathcal{H}$ and $E \subseteq K \cup P$ is finite. If some P is safe, $\mathcal{H}$ is generatable in the limit. The whole universe is safe iff no language is finitely dominated, the hypothesis of Theorem 5.2(a); in $(\mathcal{X}_0, \mathrm{cl}_0)$ the set $\{z_1, z_2, \ldots\}$ is safe iff no language has a finite core, the limit part of Proposition F.2; and the class of Example 5.5 has no safe set.*

Proof. The generator. If $\mathcal{H} \neq \emptyset$, taking $L = K$ and $E = \emptyset$ shows that a safe P meets every language; fix $p_0 \in P$. Run the scaffold of Appendix D: output

$$\mathrm{mindex}((L_{n_t} \cap P) \setminus \Psi(E_t))$$

when $\mathrm{Crit}_t \neq \emptyset$ and this set is nonempty, and p_0 otherwise.

Success. Every output lies in P and every example in the target K, so

$$E_t \subseteq K \cup P$$

at every round, including those before t^*. For $t \geq t^*$ the language L_{n_t} lies in K, and safety gives an output in $L_{n_t} \setminus \Psi(E_t) \subseteq K \setminus B_t$. No property of Ψ is used.

The instances: the whole universe. With $L = K$, safety of $\mathcal{X}$ says that no language is finitely dominated, and conversely.

The instances: master keys. In $(\mathcal{X}_0, \mathrm{cl}_0)$, if no language has a finite core, every language is infinite and holds at most one key (Proposition F.2); a finite $E \subseteq K \cup \{z_1, z_2, \ldots\}$ then holds at most one key, so $\mathrm{cl}_0(E) = E$ is finite, while $L \cap \{z_1, z_2, \ldots\}$ is infinite. Conversely, a language with a finite core makes the class not generatable in the limit (Theorem 5.2(a)), so no set is safe.

The instances: key pairs. In the key-pair space, if $a_j \in P$, take a language K that contains b_j and $L = K$: the set $E = \{a_j, b_j\} \subseteq K \cup P$ spans everything; likewise for b_j, so a safe set would be empty. □

Beyond local Noetherianity, then, what can go wrong is poisoning: a bet that is valid for one language and, together with examples of another, spans it. Bets that no examples can poison are harmless. A region fixed in advance is not necessary, however; the region may have to depend on the examples.

Example F.4 (A region chosen after the first example)**.** Let $\mathcal{X} = U \sqcup V$ with U, V infinite, let

$$\Psi(E) = \mathcal{X} \text{ if } E \text{ meets both } U \text{ and } V \qquad \text{and} \qquad \Psi(E) = E \text{ otherwise,}$$

and let $\mathcal{H} = \{\mathbf{1}_U, \mathbf{1}_V\}$. This is a finitary closure operator that is not locally Noetherian (U is dominated by any $\{u, v\}$ with $u \in U$, $v \in V$, and every finite $C \subseteq U$ spans only itself). Outputting the least unused point of the side of the first example is valid and novel against the full history from the first round: the examples and the outputs stay on the side of the target, where the span is the identity. So $\mathcal{H}$ is generatable at every level. Yet no set is safe: a safe P meets both sides (take $L = K$ and $E = \emptyset$), and for $u \in U$ and $v \in V \cap P$ the set $E = \{u, v\} \subseteq U \cup P$ spans everything. And every generator that succeeds in the limit bets only on the side of its first example: an output on the other side, together with the first example, makes the history span everything from then on, on every complete enumeration that extends that prefix.

The side of the first example is its safe zone, and the generator above is that of Theorem 4.1. So safe regions fixed in advance miss even a class of core dimension 0, while classes of infinite core dimension can have one (the whole universe, for the identity span and every class of infinite languages): Proposition F.3 and Theorem 4.1 are incomparable sufficient conditions.

Both are covered by a condition that we can state but not characterize with. *Question: is a countable class generatable in the limit, with the full history, if and only if its bets can be confined to regions $P(S)$ that grow with the sample S and such that, once S is large enough inside the target K, $L \cap P(S) \not\subseteq \Psi(E)$ for every language L with $S \subseteq L \subseteq K$ and every finite $E \subseteq S \cup P(S)$?* Fixed safe regions satisfy this condition with $P(S) = P$, and classes of finite core dimension with $P(S) = \langle S \rangle_{\mathcal{H}}$, the safe zone, which grows with S and, once S has more points than the core dimension, has no finite core.

F.8 Non-vacuity

Remark F.5 (Non-vacuity). In $(\mathbb{N}_0, \mathrm{cl}_\uparrow)$ every set is finitely dominated, so the positive side of Theorem 5.3 is vacuous there. Adjoining a free part removes this: on $\mathcal{X}' = \mathbb{N}_0 \sqcup Y$ with

$$\mathrm{cl}(A) = \mathrm{cl}_\uparrow(A \cap \mathbb{N}_0) \cup (A \cap Y),$$

the axioms hold componentwise; the space is locally Noetherian (a finitely dominated A has finite $A \cap Y$, and $A \cap Y$, together with $\min(A \cap \mathbb{N}_0)$ when $A \cap \mathbb{N}_0 \neq \emptyset$, is a finite core); BEC still fails on $\mathbb{N}_0$; and Y is not finitely dominated, as $\mathrm{cl}(F) \cap Y = F \cap Y$ is finite for finite F. So $\{\mathbf{1}_Y\}$ is generatable at every level, from the first round (its only safe zone Y has no finite core, and Theorem 4.1 applies with core dimension 0), while $\{\mathbf{1}_{\mathbb{N}_0}\}$ is generatable at none (proof of Proposition F.1).

G Matroid intersection

G.1 Proposition 7.1: capacity

Proposition 7.1 (Capacity). *If $A \subseteq \Psi(E)$ with E finite, every finite common independent $I \subseteq A$ has $|I| \leq k|E|$; if $\rho_\cap(A) < \infty$, a maximum common independent $I \subseteq A$ satisfies $A \subseteq \Psi(I)$. So A has a finite core iff it is finitely dominated iff it contains no infinite common independent set; the span is locally Noetherian, and its core dimension is $\mathrm{RCdim}_\cap \in \{\mathrm{Cdim}, \infty\}$, with $\mathrm{RCdim}_\cap \geq \max_i \mathrm{RCdim}_{M_i}$. By Theorems 4.1 and 5.2, $\mathcal{H}$ is uniformly generatable iff $\mathrm{RCdim}_\cap(\mathcal{H}) < \infty$, with threshold $\mathrm{RCdim}_\cap(\mathcal{H}) + 1$; non-uniformly iff $\mathcal{H}$ is the union of a chain of classes of finite $\mathrm{RCdim}_\cap$; and, if countable, in the limit iff every language contains an infinite common independent set.*

Proof. Per matroid $M \in \{M_1, \ldots, M_k\}$ we use:

(F0) subsets of independent sets are independent (by (CL2));
(F1) finite character: if every finite subset of I is independent, so is I, since a dependence $x \in \mathrm{cl}_M(I \setminus \{x\})$ is witnessed by a finite $F \subseteq I \setminus \{x\}$ by (CL5);
(X) if I is independent and $a \notin \mathrm{cl}_M(I)$, then $I \cup \{a\}$ is independent (the exchange step in the proof of Lemma B.3(R)).

The map Ψ is monotone and extensive, as each cl_{M_i} is.

Capacity bound. Let E be finite, $A \subseteq \Psi(E)$, and $I \subseteq A$ finite common independent. Then

$$I = \bigcup_i (I \cap \mathrm{cl}_{M_i}(E));$$

each piece is M_i-independent (F0) and contained in $\mathrm{cl}_{M_i}(E)$, so Lemma B.2 bounds it by $|E|$; summing,

$$|I| \leq k|E|.$$

Contrapositively, if $\rho_\cap(A) = \infty$ then $A \not\subseteq \Psi(E)$ for every finite E.

Maximum common independent sets span. If $\rho_\cap(A) = m < \infty$, the supremum is attained at some common independent $I \subseteq A$. If some $a \in A$ lay outside every $\mathrm{cl}_{M_i}(I)$, then $a \notin I$, and by (X) in each M_i the set $I \cup \{a\}$ would be common independent of size $m + 1$. So $A \subseteq \Psi(I)$.

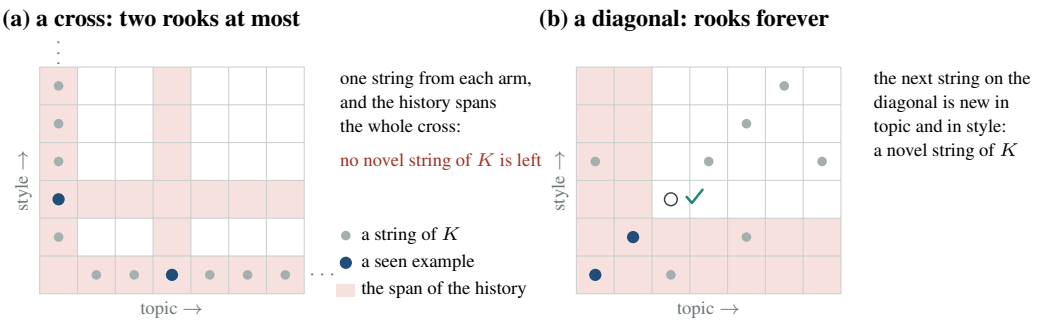

Figure 7: The example of Proposition 7.2 on the grid of topics and styles. A common independent set is a placement of non-attacking rooks, and the span of the history (red) is the union of the columns and rows of its strings. (a) The cross, one column and one row, reaches infinitely many topics and infinitely many styles, but once one string from each arm has been seen the spans of the history cover it. (b) A language containing the diagonal holds infinitely many non-attacking rooks, and its one-language class is generatable in the limit.

Infinite capacity. If A contains an infinite common independent J, its finite subsets, which are common independent by (F0), witness $\rho_\cap(A) = \infty$. Conversely, if $\rho_\cap(A) = \infty$, the greedy recursion

$$j_1 = \mathrm{mindex}(A \setminus \Psi(\emptyset)), \qquad j_{m+1} = \mathrm{mindex}(A \setminus \Psi(\{j_1, \ldots, j_m\}))$$

is total by the capacity bound; as $j_{m+1} \notin \mathrm{cl}_{M_i}(\{j_1, \ldots, j_m\})$ for every i, each prefix is M_i-independent for every i by Lemma B.1, and the union is common independent by (F0) and (F1). For $k = 1$, $\rho_\cap = r_M$ by Lemma B.3(R).

Cores and local Noetherianity. A finite core makes A finitely dominated; if A is finitely dominated, the capacity bound gives $\rho_\cap(A) < \infty$; and then a maximum common independent subset of A, which is finite, is a finite core. So the three conditions of the statement coincide, and the span is locally Noetherian.

The core dimension. The witness condition of $\mathrm{Coredim}_\Psi$ is therefore that the safe zone has finite capacity, which is the witness condition of $\mathrm{RCdim}_\cap$; so $\mathrm{Coredim}_\Psi = \mathrm{RCdim}_\cap$, and $\mathrm{RCdim}_\cap \in \{\mathrm{Cdim}, \infty\}$ by Theorem 4.1. A common independent set is M_i-independent, so $\rho_\cap \le r_{M_i}$ by Lemma B.3(R); hence an RCdim_{M_i} witness is an $\mathrm{RCdim}_\cap$ witness, and $\mathrm{RCdim}_\cap \ge \max_i \mathrm{RCdim}_{M_i}$.

Generation. Theorem 4.1, with $\mathrm{Coredim}_\Psi = \mathrm{RCdim}_\cap$, gives the uniform characterization and the threshold. The span is locally Noetherian, so Theorem 5.2(b) characterizes non-uniform generatability by chains of classes of finite $\mathrm{Coredim}_\Psi = \mathrm{RCdim}_\cap$, and Theorem 5.2(a) characterizes generatability in the limit of a countable class by the absence of a language with a finite core, that is, by every language containing an infinite common independent set. Finiteness of k enters only through the factor k in the capacity bound. □

G.2 Proposition 7.2: Rank in each matroid is not enough

Proposition 7.2 (Rank in each matroid is not enough)**.** *There are finitary matroids M_1, M_2 on a countable $\mathcal{X}$ and a one-language class $\mathcal{H} = \{h\}$ with $r_{M_1}(\mathrm{supp}\, h) = r_{M_2}(\mathrm{supp}\, h) = \infty$ and $\mathrm{supp}(h) = I_1 \cup I_2$, where I_i is independent in M_i, such that nevertheless $\rho_\cap(\mathrm{supp}\, h) = 2$ and $\mathcal{H}$ is not simultaneously generatable at any level.*

Proof. The matroids and the class. (See Figure 7.) Let $\mathcal{X} = P \sqcup Q$ with $P = \{a_1, a_2, \ldots\}$ and $Q = \{b_1, b_2, \ldots\}$, and define

$$\mathrm{cl}_{M_1}(S) = S \text{ if } S \cap Q = \emptyset, \text{ else } S \cup Q$$

(free on P, a single parallel class on Q), and M_2 the mirror. In the language of topics and styles, a_i has topic i and style 0, and b_j has topic 0 and style j. Each M_i is a finitary matroid, the direct sum of a free matroid and a rank-one uniform matroid on an infinite set. Take $\mathcal{H} = \{h\}$ with $\mathrm{supp}(h) = \mathcal{X}$.

Infinite rank in each matroid. The set $\{a_1, \ldots, a_m\}$ is M_1-independent for every m, so $r_{M_1}(\mathcal{X}) = \infty$ by Lemma B.3(R), and symmetrically for M_2; and $\mathcal{X} = P \cup Q$ with P independent in M_1 and Q independent in M_2.

Capacity two. A common independent J has $|J \cap Q| \leq 1$ (two points of Q are M_1-dependent) and $|J \cap P| \leq 1$, so $\rho_\cap(\mathcal{X}) = 2$, attained by $\{a_1, b_1\}$.

Generatable at no level. So the language contains no infinite common independent set, and every set of distinct points is consistent with safe zone $\mathcal{X}$, of finite capacity: $\mathrm{RCdim}_\cap(\mathcal{H}) = \infty$, and every chain with union $\mathcal{H}$ has $\mathcal{H}$ itself as a member. By Proposition 7.1, $\mathcal{H}$ is generatable at no level. Directly: once a_1 and b_1 have both appeared, every later history meets P and Q, so $\Psi(E_s) \supseteq (E_s \cup Q) \cup (E_s \cup P) = \mathcal{X}$. □

H Extended related work

H.1 Metric novelty and covering centres

Novelty relations and covering centres. Li et al. (2026a) study metric novelty in separable metric spaces: an output must avoid the closed ε'-balls around the adversary's examples, an examples-only span, and the count of examples is replaced by an ε-covering number. Their dimension asks for safe zones covered by finitely many ε'-balls centred anywhere in the space (Li et al., 2026a, Defs. 2.2 and 3.1, arXiv v1). The centres matter.

Let $\mathcal{X} \subset \ell^2$ consist of 0, the points $e_1, e_2, \ldots$ and $10f_1, 10f_2, \ldots$, with all e_j and f_k jointly orthonormal, and let

$$\mathrm{supp}(h_n) = \{e_k : k \geq 1\} \cup \{10f_k : k \geq n\}.$$

Distinct points of the languages are at least $\sqrt{2}$ apart. So for $\varepsilon \leq 1/2$ the ε-covering number of a finite set of examples is its size, and for the ε'-neighbourhood span with $\varepsilon' = 6/5$ no safe zone has a finite core: outputting the least unseen e_k is valid and novel from the first example, as Proposition 5.1 predicts. Yet the safe zone of $e_1, \ldots, e_d$ is $\{e_k : k \geq 1\}$, which lies in the ε'-ball about 0, a point of no language. The points $10f_k$ are $10\sqrt{2} > 4$ apart, so every language has infinite 2-covering number, as their Def. 2.3 requires with $r = 2$. With $(\varepsilon, \varepsilon', r) = (1/2, 6/5, 2)$ the class is therefore uniformly generatable while their dimension is infinite: the only-if direction of their Theorem 3.1 fails as stated.

The covering step in the proof of their Lemma B.1 takes the centres in the safe zone. A cover centred in the set is a finite core. Beyond this, our novelty can also bind against the generator's own outputs, and it is set-wise rather than pairwise.

H.2 Other work

Breadth and diversity. Consistency-versus-breadth trade-offs and their characterizations appear in Kalavasis et al. (2025; 2024); Charikar & Pabbaraju (2025); output density in Kleinberg & Wei (2025a;b; 2026); a γ-fraction novelty dial in Strauss et al. (2026); representativeness across groups, with a group closure dimension, in Peale et al. (2025). Rank novelty differs from all of these in constraining the internal structure of the output set rather than its coverage of K. In practice, diversity is often scored by the spread of feature vectors, through determinantal point processes (Kulesza & Taskar, 2012) or the Vendi score (Friedman & Dieng, 2023); rank novelty is a worst-case, in-the-limit counterpart of such scores, not a proposal to replace them.

Arbitrary classes. For plain novelty, generation in the limit is characterized for arbitrary classes over a countable universe by finite witnesses (Li et al., 2026b); our limit-level results are for countable classes, and whether cores can play the role of finite witnesses for spans is open.

Variants of the protocol. Noise, loss and feedback (Bai et al., 2026; Li & Zhang, 2026), contaminated or private streams (Mehrotra et al., 2025; 2026), memory and space constraints (Kleinberg et al., 2026a; Flammarion et al., 2026), mistake bounds (Kleinberg et al., 2026b), safety constraints (Anastasopoulos et al., 2026), contrastive data (Li et al., 2026c), time sensitivity (Ganju et al., 2026), replay (Racca et al., 2026), and computational barriers (Arenas et al., 2025) modify the resources or the target; among these works none modifies the novelty relation or places algebraic structure on the universe.

Union-closure of generatability is studied by Hanneke et al. (2025); our union-versus-intersection phenomenon concerns unions of *matroids*, an orthogonal question.

Prompted generation was introduced by Kleinberg & Mullainathan (2024) and characterized by Raman et al. (2025); we do not touch it.

Matroids in learning. Identification-type learning of closed sets in matroids and closure systems was studied by Gao et al. (2012); de Brecht et al. (2007); matroid oracle models appear in Coullard & Hellerstein (1996); closure queries in exact learning in Arias et al. (2015). None concerns generation. Hallucination detection (Karbasi et al., 2025) and offline diverse selection under quotas (Moumoulidou et al., 2020) are adjacent in motivation only.