Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 62 (2026)

Toward a Characterization of Simulation Between Arithmetic Theories

Hunter Monroe

July 2026

Abstract

We study when a sound arithmetic theory $\mathcal{S}\supseteq S^1_2$ with polynomial-time decidable axioms efficiently proves the bounded consistency statements $Con_{\mathcal{S}+\phi}(n)$ for a true sentence ϕ. Equivalently, we ask when $\mathcal{S}$, viewed as a proof system, simulates $\mathcal{S}+\phi$. The paper's two unconditional contributions constrain possible characterizations. First, for finitely axiomatized sequential $\mathcal{S}$, if $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+\phi}$, then $\mathcal{S}$ interprets $\mathcal{S}+\phi$, implying $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}}(p(n))\rightarrow Con_{\mathcal{S}+\phi}(n)$ for some polynomial p, and hence $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$. Second, if $\mathcal{S}$ fails to simulate $\mathcal{S}+\phi$ for some true ϕ, then for all sufficiently large k it also fails to simulate $S^1_2+\phi_{BB}(k)$, where $\phi_{BB}(k)$ asserts the exact value of the k-state Busy Beaver function. Informally, any hard true extension yields a canonical Busy Beaver witness to the same obstruction. These results suggest that relative consistency strength is a serious candidate for governing when simulation is possible, while leaving open whether it is the correct criterion.

The paper's central conjectural proposal is: for finitely axiomatized sequential $\mathcal{S}$, if $EA\nvdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+\phi}$, then for every constant $c>0$, $\mathcal{S}\nvdash^{n^c} Con_{\mathcal{S}+\phi}(n)$. Under this proposal, hardness follows in canonical cases where ϕ is $Con_{\mathcal{S}}$ or a Kolmogorov-randomness axiom. The latter yields further conjectural consequences and extensions.

1 Introduction

"...the essence of the big open problems in complexity theory could be logical, rather than combinatorial" — Pudlák [13]

ISSN 1433-8092

Let $\mathcal{S}\supseteq S^1_2$ be a sound arithmetic theory with polynomial-time decidable axioms, and let ϕ be a true sentence independent of $\mathcal{S}$. We study when $\mathcal{S}$ efficiently proves the bounded consistency statements $Con_{\mathcal{S}+\phi}(n)$, where $Con_{\mathcal{S}+\phi}(n)$ asserts that $\mathcal{S}+\phi$ has no proof of 0=1 within n symbols.

From the perspective of proof complexity, this asks when $\mathcal{S}$, viewed as a proof system, simulates $\mathcal{S}+\phi$.[1] Independence alone does not prevent simulation: there are true independent sentences ϕ such that $\mathcal{S}$ polynomial-time interprets $\mathcal{S}+\phi$, and hence efficiently proves $Con_{\mathcal{S}+\phi}(n)$; see Pudlák [14, Lemma 3.6]. On the other hand, if no optimal proof system exists, then for some true sentence ϕ one has for all c, $\mathcal{S}\not\vdash^{n^c} Con_{\mathcal{S}+\phi}(n)$; see Krajíček and Pudlák [8, Theorem 2.1].[2] Thus the problem is to identify a criterion for ϕ that governs simulation and non-simulation.

We formulate this as follows.

> **Characterization Problem.** For every sound theory $\mathcal{S}\supseteq S^1_2$ and every true sentence ϕ independent of $\mathcal{S}$, determine when $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$, and when instead $\mathcal{S}\not\vdash^{n^c} Con_{\mathcal{S}+\phi}(n)$ for every constant c. Equivalently, determine when $\mathcal{S}$ simulates $\mathcal{S}+\phi$ as a proof system.[3]

A benchmark case is $\phi{=}Con_{\mathcal{S}}$; Pudlák [12, Problem 1] conjectures that $\mathcal{S}\not\vdash^{n^c} Con_{\mathcal{S}+Con_{\mathcal{S}}}(n)$ for every c. This conjecture implies that higher absolute consistency strength is a sufficient condition for non-simulation (Theorem 3.6 below). The present paper asks for a more general organizing principle for arbitrary true independent sentences ϕ, ideally a condition that is both necessary and sufficient.

The paper's first contribution identifies a new sufficient condition for simulation for finitely axiomatized sequential $\mathcal{S}$. If $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+\phi}$, then $\mathcal{S}$ interprets $\mathcal{S}+\phi$, where EA is Elementary Arithmetic (Visser [16]). By proof

[1] For background on proof complexity, see Krajíček[7].

[2] We use the notation $\mathcal{S}\vdash^{n^{O(1)}}\phi(n)$ to mean that there exists a constant c such that $\mathcal{S}$ has proofs of $\phi(n)$ of size at most n^c for all sufficiently large n. Hardness statements are written explicitly in the form "for every constant c, $\mathcal{S}\not\vdash^{n^c}\phi(n)$."

[3] We limit attention to true ϕ to isolate non-simulation arising from lack of access to additional *correct* information, rather than trivial failure due to unsound extensions, and because this is the regime relevant for applications to tautologies, where ϕ expresses the truth of a family of propositional formulas.

translation, there exists a polynomial p such that $\mathcal{S}\vert\!\frac{n^{O(1)}}{}\, Con_{\mathcal{S}}(p(n)) \to Con_{\mathcal{S}+\phi}(n)$. Thus Theorem 3.4 shows that relative consistency ($EA \vdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+\phi}$) implies simulation ($\mathcal{S}\vert\!\frac{n^{O(1)}}{}\, Con_{\mathcal{S}+\phi}(n)$) for $\mathcal{S}$ finitely axiomatized and sequential. Taking this theorem to be tight leads to the following structural candidate characterization of simulation. The paper develops that proposal and its main consequences. In particular, the case $\phi = Con_{\mathcal{S}}$ and the Kolmogorov-randomness axioms arise directly as natural instances of the same general obstruction.

The paper's second contribution shows that canonical incompleteness phenomena already suffice to witness non-simulation. More precisely, if $\mathcal{S}$ fails to simulate $\mathcal{S}+\phi$ for some true sentence ϕ, then for all sufficiently large k it also fails to simulate $S^1_2 + \phi_{BB}(k)$, where $\phi_{BB}(k)$ asserts the exact value of the k-state Busy Beaver function, namely the maximum halting time of any halting k-state Turing machine (Theorem 3.10).[4] The change in base theory here is deliberate: the theorem shows that once hardness occurs above S^1_2, it is already witnessed by a canonical Busy Beaver extension of the weak base theory S^1_2. This aligns with Aaronson's observation that, for sufficiently large k, true Busy Beaver sentences can prove the consistency of arbitrarily strong computably axiomatized theories [1, Proposition 3]. Thus the hardness side of the Characterization Problem already appears among Busy Beaver sentences. Any characterization of simulation should therefore explain these canonical unprovable truths.

Together, these results constrain possible characterizations and suggest two complementary ways of thinking about hardness. First, in line with the Busy Beaver reduction, efficient simulation should be controlled by information already visible to weak arithmetic; this points toward a structural criterion formulated in terms of relative-consistency transfer. Second, canonical sources of unprovable information such as Busy Beaver values and Kolmogorov-random strings suggest a more semantic or information-theoretic obstruction: simulation should fail when it would require access to true information that the base theory cannot itself recover in any way certified in a weak base theory, such as EA. Although weak, such a theory can prove true statements about Turing machines that halt.

These considerations motivate a single structural picture of hardness. The main proposal of the paper is **Higher Relative Consistency (HRC)**, which asserts that, for a true sentence ϕ, efficient simulation fails exactly when weak

[4] See Rado [15].

arithmetic cannot certify that adjoining ϕ preserves consistency.

> **Higher Relative Consistency (informal).** For every true sentence ϕ: if $EA \nvdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+\phi}$, then for every constant c, $\mathcal{S} \not\vdash^{n^c} Con_{\mathcal{S}+\phi}(n)$.

In other words, failure of relative-consistency should already rule out efficient simulation. Combined with the positive simulation theorem, this yields a natural conjectural picture in which simulation occurs exactly when weak arithmetic can certify that adjoining ϕ preserves consistency.

The remainder of the paper is organized as follows. Section 2 provides preliminaries. Section 3 presents unconditional constraints on any characterization. Section 4 develops heuristic support for the proposed criterion. Section 5 states **Higher Relative Consistency**, derives its main consequences, and formulates *Kolmogorov Hardness* as a distinguished random-axiom instance of the general characterization. Section 6 develops extensions of Kolmogorov Hardness. Section 7 discusses provability of **Higher Relative Consistency**. Section 8 concludes.

2 Preliminaries

We work with sound arithmetical theories $\mathcal{S} \supseteq S^1_2$ whose axioms are decidable in polynomial time. Throughout, ϕ denotes a true arithmetical sentence, and $\mathcal{S}+\phi$ is the corresponding true extension of $\mathcal{S}$. Unless explicitly stated otherwise, implications of the form $Con_{\mathcal{S}} \to Con_{\mathcal{S}+\phi}$ are understood as formalized in EA. This is the level at which the interpretability criteria used later are naturally stated.

We fix a standard arithmetization of syntax and a fixed presentation of each theory under discussion. Throughout, $Con_{\mathcal{S}}$ denotes the corresponding arithmetized consistency predicate. Proofs are encoded as binary strings, and all proof lengths are measured in symbols under this encoding. As usual, any two reasonable codings yield polynomially equivalent proof lengths, so statements of the form $\mathcal{S} \vdash^{n^{O(1)}} \phi(n)$ are robust under the choice of coding.

For a theory $\mathcal{S}$, let $Con_{\mathcal{S}}(n)$ denote the bounded consistency statement asserting that there is no $\mathcal{S}$-proof of $0{=}1$ of length at most n. We write $\mathcal{S} \vdash^{n^c} \phi(n)$ to mean that, for all sufficiently large n, the sentence $\phi(n)$ has an

$\mathcal{S}$-proof of length at most n^c. Likewise, $\mathcal{S}\overset{n^{O(1)}}{\vdash}\phi(n)$ means that such proofs exist with polynomially bounded length.

We say that $\mathcal{S}$ *simulates* $\mathcal{S}+\phi$ if $\mathcal{S}\overset{n^{O(1)}}{\vdash}Con_{\mathcal{S}+\phi}(n)$. Thus the Characterization Problem asks when $\mathcal{S}$ admits polynomial-size proofs of the bounded consistency statements for the true extension $\mathcal{S}+\phi$.

We say that ϕ *raises the relative-consistency strength* of $\mathcal{S}$ if $EA\nvdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$, where EA is Elementary Arithmetic (Visser [16]). We introduce new terminology, saying that $\mathcal{S}$ has *feasible relative consistency* for ϕ if there is a polynomial p such that $\mathcal{S}\overset{n^{O(1)}}{\vdash}Con_{\mathcal{S}}(p(n))\to Con_{\mathcal{S}+\phi}(n)$.

We distinguish throughout between external proof-length bounds and formalized internal implications. The statement $\mathcal{S}\overset{n^{O(1)}}{\vdash}Con_{\mathcal{S}+\phi}(n)$ is an external assertion about the existence of short $\mathcal{S}$-proofs, whereas $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$ is an internal relative-consistency implication in a weak base theory. Much of the paper is concerned with when the former should force the latter.

Finally, we use standard notions of interpretability. When additional hypotheses such as finite axiomatizability or sequentiality are needed, they will be stated explicitly. In the finitely axiomatized sequential setting used below, the relevant equivalence is the Friedman–Visser interpretability criterion: Friedman's characterization treats the appropriate consistency statement for a finitely axiomatized target as the weakest sentence whose adjunction yields an interpretation, and Visser's Interpretation Existence Lemma supplies the weak-base relative-consistency-to-interpretation direction [16]. Thus, in this regime, $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$ is equivalent to an interpretation of $\mathcal{S}+\phi$ in $\mathcal{S}$, relative to the fixed presentations and formalized consistency predicates. Polynomial-time interpretation yields polynomial-overhead proof translation. These standard facts underlie the paper's unconditional upper-bound mechanism.

Whenever an external proof transformation is used to derive polynomial-size $\mathcal{S}$-proofs, we will state explicitly whether the transformation is merely true in the standard model, formalizable in EA, or available with polynomial-size $\mathcal{S}$-proofs.

These notions fix the framework for the paper's central question: whether simulation of $\mathcal{S}+\phi$ by $\mathcal{S}$ is governed by relative-consistency transfer, and in particular by the weak-base implication $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$.

3 Unconditional Constraints on Characterizations

This section presents the paper's unconditional results. The first shows in certain settings that relative consistency implies feasible relative consistency and hence simulation. The second shows that any hard true ϕ can, in a precise sense, be replaced by one of Busy Beaver form. Together, these results isolate both the positive mechanism underlying simulation and a canonical source of hardness, constraining any possible characterization.

The literature shows that simulation is possible in special cases. In particular, polynomial-time interpretability implies efficient provability of the bounded consistency statements $Con_{\mathcal{S}+\phi}(n)$; see Theorem 3.1.[5] By contrast, Pudlák [12] conjectures that this fails already for $\phi{=}Con_{\mathcal{S}}$.[6] Unconditional lower bounds of the form $\mathcal{S} \mathrel{|\!\!\!\overset{n^{O(1)}}{\not\!\!\!-\!\!-}} Con_{\mathcal{S}+\phi}(n)$ are not known for sound theories $\mathcal{S}{\supseteq}S^1_2$; such a result would imply that no optimal proof system exists, hence $\mathbf{NP}{\neq}\mathbf{coNP}$.

The two results proved in this section constrain possible answers to the Characterization Problem in complementary ways. The first isolates a sufficient condition for simulation, formulated in terms of relative consistency rather than polynomial-time interpretability. The second shows that any instance of non-simulation can already be witnessed by ϕ referring to Busy Beavers. Together, these results clarify both the positive mechanism underlying simulation and a canonical source of hardness.

3.1 A Known Sufficient Condition for Simulation

A proof translation argument by Jeřábek, as presented by Pudlák [14, Lemma 3.6], shows that if $\mathcal{S}$ polynomial-time interprets $\mathcal{S}{+}\phi$, then polynomial-size proofs of $Con_{\mathcal{S}}(n)$ yield polynomial-size proofs of $Con_{\mathcal{S}+\phi}(n)$. Moreover, for a given $\mathcal{S}$, there are true sentences ϕ independent of $\mathcal{S}$ such that $\mathcal{S}$

[5] Freund and Pakhomov [4] prove polynomial-size finite-consistency proofs for a slow-consistency extension. This is a calibration test, not by itself a counterexample to the proposed characterization: short bounded proofs do not determine whether the exact global EA relative-consistency implication used here holds for the chosen presentations, and the infinitely axiomatized setting requires separate interpretability analysis.

[6] Khaniki [6] studies computable jump operators, which produce from every proof system one that it cannot simulate; whether such operators exist is open, and Pudlák's Conjecture would supply one via $\mathcal{S}{\mapsto}\mathcal{S}{+}Con_{\mathcal{S}}$.

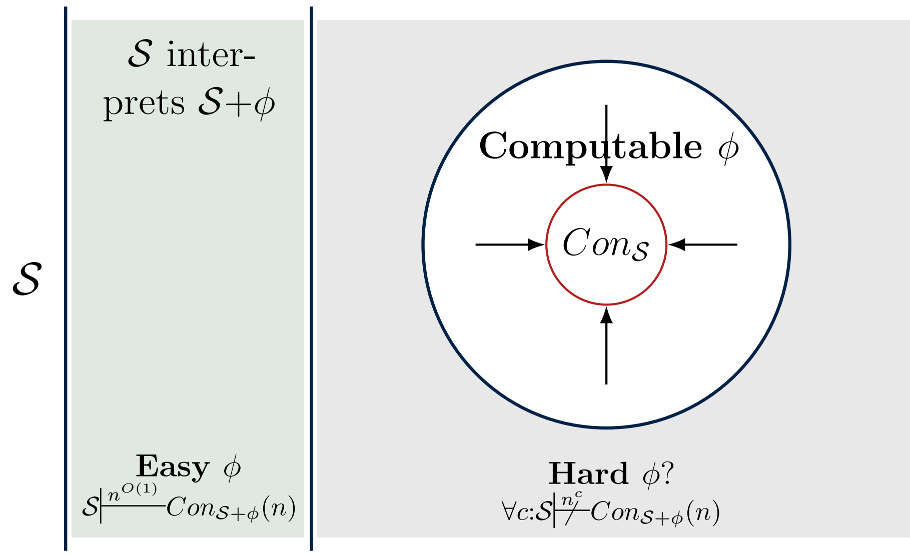

Figure 1: Current knowledge about simulation by $\mathcal{S}$. The left region records the known easy zone: when $\mathcal{S}$ interprets $\mathcal{S}+\phi$, one obtains feasible bounded-consistency proofs [14]. The middle region records the open problem for computable true ϕ, with $Con_{\mathcal{S}}$ as the central test case. Khaniki [6] shows that if a computable jump operator exists, then Pudlák's conjecture holds; equivalently, computable hard-jump behavior transfers to the canonical consistency extension $Con_{\mathcal{S}}$.

polynomial-time interprets $\mathcal{S}+\phi$, for example H-Rosser sentences as in Hájek–Pudlák [5, Theorem 4.5(5)]. Thus, simulation is not ruled out for every true independent sentence ϕ—non-simulation requires more than independence. Figure 1 summarizes current knowledge.

We present that argument with a theorem statement and proof in a form convenient for this paper's exposition.

Theorem 3.1. *(Jeřábek, via Pudlák [14, Lemma 3.6]) Suppose* $\mathcal{S}\supseteq S^1_2$ *and* $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}}(n)$*. If* $\mathcal{S}$ *polynomial-time interprets* $\mathcal{S}+\phi$*, then* $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$*.*

Proof. Let i be a polynomial-time interpretation of $\mathcal{S}+\phi$ in $\mathcal{S}$. By the standard proof-translation argument for interpretations, as presented in Pudlák [14, Lemma 3.6], there is a polynomial p such that from every $(\mathcal{S}+\phi)$-proof π of a sentence ψ one can compute an $\mathcal{S}$-proof of the translated sentence ψ^i of length at most $p(|\pi|)$.

Apply this to $\psi \equiv 0{=}1$. Then any $(\mathcal{S}{+}\phi)$-proof of contradiction of length at most n yields an $\mathcal{S}$-proof of $(0{=}1)^i$ of length at most $p(n)$. Since i is an interpretation, $(0{=}1)^i$ is a fixed false sentence, and there is a constant-size $\mathcal{S}$-proof of $(0{=}1)^i{\to}0{=}1$. After increasing p if necessary, it follows that any $(\mathcal{S}{+}\phi)$-proof of contradiction of length at most n yields an $\mathcal{S}$-proof of $0{=}1$ of length at most $p(n)$.

The proof translation and its polynomial bound are formalizable in S^1_2, hence in $\mathcal{S}$. Therefore $\mathcal{S}$ proves that if there is no $\mathcal{S}$-proof of contradiction of length at most $p(n)$, then there is no $(\mathcal{S}{+}\phi)$-proof of contradiction of length at most n. That is, $\mathcal{S}{\vdash}\forall n{:}Con_{\mathcal{S}}(p(n)){\to}Con_{\mathcal{S}+\phi}(n)$.

By assumption, $\mathcal{S}\overset{n^{O(1)}}{\vdash}Con_{\mathcal{S}}(n)$. Substituting $p(n)$ for n gives polynomial-size $\mathcal{S}$-proofs of $Con_{\mathcal{S}}(p(n))$. Composing these with this implication yields polynomial-size $\mathcal{S}$-proofs of $Con_{\mathcal{S}+\phi}(n)$. Hence $\mathcal{S}\overset{n^{O(1)}}{\vdash}Con_{\mathcal{S}+\phi}(n)$. $\square$

The essential content of the interpretation argument is not merely simulation, but the existence of a uniform feasible relative-consistency implication from $\mathcal{S}$ to $\mathcal{S}{+}\phi$. All currently known positive mechanisms for simulation pass through such implications.

A central question is the converse: whether every instance of simulation must admit such an internal explanation.

3.2 Stronger Results on Simulation

The proof of Theorem 3.1 relies only on the existence of a polynomial p such that $\mathcal{S}{\vdash}\forall n{:}Con_{\mathcal{S}}(p(n)) \to Con_{\mathcal{S}+\phi}(n)$. This suggests the following sufficient condition for simulation: $\mathcal{S}\overset{n^{O(1)}}{\vdash}Con_{\mathcal{S}}(p(n)) \to Con_{\mathcal{S}+\phi}(n)$. We call this *feasible relative consistency.*

Theorem 3.2. *Let $\mathcal{S}{\supseteq}S^1_2$ be a sound theory with polynomial-time decidable axioms such that $\mathcal{S}\overset{n^{O(1)}}{\vdash}Con_{\mathcal{S}}(n)$. Then the following are equivalent:*

1. $\mathcal{S}\overset{n^{O(1)}}{\vdash}Con_{\mathcal{S}}(p(n)){\to}Con_{\mathcal{S}+\phi}(n)$ *for some polynomial p.*

2. $\mathcal{S}\overset{n^{O(1)}}{\vdash}Con_{\mathcal{S}+\phi}(n)$.

Proof. For (1)→(2), combine polynomial-size proofs of $Con_{\mathcal{S}}(p(n))$ with polynomial-size proofs of $Con_{\mathcal{S}}(p(n)){\to}Con_{\mathcal{S}+\phi}(n)$. This yields polynomial-size proofs of $Con_{\mathcal{S}+\phi}(n)$.

For (2)→(1), assume $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$. Fix any polynomial p. From a proof of $Con_{\mathcal{S}+\phi}(n)$, one can derive a proof of $Con_{\mathcal{S}}(p(n))\to Con_{\mathcal{S}+\phi}(n)$ with only polynomial overhead by prefixing a fixed implicational derivation. Since the formula $Con_{\mathcal{S}}(p(n))$ has size polynomial in n, the resulting proof length remains polynomially bounded. Hence $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}}(p(n))\to Con_{\mathcal{S}+\phi}(n)$. □

Accordingly, in what follows we treat simulation itself, namely $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$, as the primary notion. When $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}}(n)$, feasible relative consistency is simply an equivalent reformulation: condition (1) expresses feasible relative consistency, and condition (2) expresses simulation. To state the contrapositive of the above theorem:

Corollary 3.3. *Let $\mathcal{S}\supseteq S^1_2$ be a sound theory with polynomial-time decidable axioms such that $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}}(n)$. Then, for any polynomial p, the following are equivalent:*

1. $\mathcal{S} \not\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$.
2. $\mathcal{S} \not\vdash^{n^{O(1)}} Con_{\mathcal{S}}(p(n))\to Con_{\mathcal{S}+\phi}(n)$.

Proof. This follows immediately from Theorem 3.2 by contraposition in both directions. □

The interpretation theorem yields the following stronger positive result in the finitely axiomatized sequential setting.

Theorem 3.4. *Suppose $\mathcal{S}\supseteq S^1_2$ is finitely axiomatized and sequential. If $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$, then there exists a polynomial p such that $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}}(p(n))\to Con_{\mathcal{S}+\phi}(n)$. In particular, $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$.*

Proof. Pudlák [12] shows $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}}(n)$. By the Friedman–Visser interpretability criterion for finitely axiomatized sequential theories, with the relative-consistency-to-interpretation direction supplied by Visser's Interpretation Existence Lemma [16], the hypothesis $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$ implies that $\mathcal{S}$ interprets $\mathcal{S}+\phi$.

By Theorem 3.1, it follows that $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$. The existence of a polynomial p such that $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}}(p(n))\to Con_{\mathcal{S}+\phi}(n)$ then follows from Theorem 3.2. □

By contrast, Pudlák [12] conjectures that even the benchmark case $\phi{=}Con_{\mathcal{S}}$ already yields non-simulation:

Conjecture 3.5. *(Pudlák's Conjecture) For every constant c, $\mathcal{S}\mathrel{\not|\!\!\frac{n^c}{}} Con_{\mathcal{S}+Con_{\mathcal{S}}}(n)$.*[7]

Under Pudlák's Conjecture, any true extension that raises absolute consistency (proves $Con_{\mathcal{S}}$) is hard:

Theorem 3.6. *Assume Pudlák's Conjecture, and suppose $\mathcal{S}{+}\phi{\vdash}Con_{\mathcal{S}}$. Then for every constant c, $\mathcal{S}\mathrel{\not|\!\!\frac{n^c}{}} Con_{\mathcal{S}+\phi}(n)$.*

Proof. Fix a proof π of $Con_{\mathcal{S}}$ in $\mathcal{S}{+}\phi$.

There is a uniform proof transformation sending any $(\mathcal{S}{+}Con_{\mathcal{S}})$-proof to an $(\mathcal{S}{+}\phi)$-proof by replacing each use of the extra axiom $Con_{\mathcal{S}}$ by the fixed derivation π. This increases proof length by at most a constant multiplicative and additive factor. Hence there exist a linear polynomial p and a constant d such that $\mathcal{S}\mathrel{|\!\!\frac{n^d}{}}Con_{\mathcal{S}+\phi}(p(n)){\to}Con_{\mathcal{S}+Con_{\mathcal{S}}}(n)$.

Suppose toward a contradiction that $\mathcal{S}\mathrel{|\!\!\frac{n^{O(1)}}{}}Con_{\mathcal{S}+\phi}(n)$. Since p is linear, substituting $p(n)$ for n still gives $\mathcal{S}\mathrel{|\!\!\frac{n^{O(1)}}{}}Con_{\mathcal{S}+\phi}(p(n))$. Composing these proofs with this implication yields $\mathcal{S}\mathrel{|\!\!\frac{n^{O(1)}}{}}Con_{\mathcal{S}+Con_{\mathcal{S}}}(n)$, contrary to Pudlák's Conjecture.

Therefore for every constant c, $\mathcal{S}\mathrel{\not|\!\!\frac{n^c}{}}Con_{\mathcal{S}+\phi}(n)$. ☐

Then there is a gap between Pudlák's Conjecture and the strongest known sufficient condition for simulation, in Theorem 3.4:

Corollary 3.7. *Pudlák's Conjecture is not tight as a converse to Theorem 3.4.*

Proof. Let θ be any true sentence such that $\mathcal{S}{+}Con_{\mathcal{S}}{\nvdash}\theta$, and put $\phi{:=}Con_{\mathcal{S}}{\wedge}\theta$. Then $\mathcal{S}{+}\phi{\vdash}Con_{\mathcal{S}}$, so by Theorem 3.6, one has, for every constant c, $\mathcal{S}\mathrel{\not|\!\!\frac{n^c}{}} Con_{\mathcal{S}+\phi}(n)$.

However, ϕ is strictly stronger than $Con_{\mathcal{S}}$ over $\mathcal{S}$. Indeed, $\mathcal{S}{\vdash}\phi{\to}Con_{\mathcal{S}}$ is immediate. If also $\mathcal{S}{\vdash}Con_{\mathcal{S}}{\to}\phi$, then $\mathcal{S}{+}Con_{\mathcal{S}}{\vdash}\phi$, hence $\mathcal{S}{+}Con_{\mathcal{S}}{\vdash}\theta$, contrary to the choice of θ.

[7]Another natural candidate for a hard extension is given by the Buss jump [2], that is, a sentence or schema asserting the soundness of the next level of bounded reasoning over $\mathcal{S}$. If a chosen formalization yields a true sentence ϕ_{Jump} such that $\mathcal{S}{+}\phi_{\text{Jump}} \vdash Con_{\mathcal{S}}$, then Pudlák's Conjecture already implies that $\mathcal{S}$ does not simulate $\mathcal{S}{+}\phi_{\text{Jump}}$. The stronger **Higher Relative Consistency** assumption yields the same conclusion directly.

Thus, the same hardness conclusion holds not only for the extension $\mathcal{S}+Con_{\mathcal{S}}$, but also for strictly stronger true extensions $\mathcal{S}+\phi$. Therefore Pudlák's Conjecture is not tight as a converse to Theorem 3.4. $\square$

Thus, there is a gap between Theorem 3.4 and Pudlák's Conjecture. The former is governed by feasible relative consistency, while the latter already yields hardness for extensions $\mathcal{S}+\phi$ satisfying $\mathcal{S}+\phi\vdash Con_{\mathcal{S}}$, that raise absolute consistency. A central aim of the paper is to isolate principles that close this gap.

3.3 Non-Simulation Implies Non-Simulation for Busy Beavers

A theory $\mathcal{S}$ for which, for every true sentence ϕ, there exists a polynomial bound on $\mathcal{S}$-proofs of $Con_{\mathcal{S}+\phi}(n)$—would seem to support simulations that cannot be explained solely by information provable in $\mathcal{S}$, since $\mathcal{S}$ does not prove all true sentences. We show that this phenomenon is already witnessed by a canonical family of true sentences, namely exact Busy Beaver value statements.

For each k, let $t_k=BB(k)$ be the true k-state Busy Beaver value, and let $\phi_{BB}(k)$ denote the true sentence asserting the exact value $BB(k)=t_k$. Since t_k is the maximum halting time of any halting k-state machine on blank input, S^1_2 proves that $\phi_{BB}(k)$ implies that every halting k-state machine on blank input halts within t_k steps. We will use this bounded-halting consequence in the proof of Lemma 3.9.

We begin by proving that, for any fixed true computably axiomatized theory, sufficiently large Busy Beaver axioms already imply its consistency (Aaronson [1, Proposition 3]). The relevant threshold is the size needed to realize the contradiction-search machine for the theory. Once this is in place, any hard true extension of a sound theory $\mathcal{S}\supseteq S^1_2$ yields eventual hardness throughout the Busy Beaver family.

Definition 3.8. (**Contradiction Search Threshold**) Let $\mathcal{T}$ be a computably axiomatized theory, and let $E_{\mathcal{T}}$ be a Turing machine which enumerates $\mathcal{T}$-proofs and halts exactly when it finds a proof of contradiction. Define $k^{\mathrm{cs}}_{\mathcal{T}}$ to be any threshold such that for every $k\geq k^{\mathrm{cs}}_{\mathcal{T}}$, there is a k-state Turing machine computing the same partial function as $E_{\mathcal{T}}$.

Lemma 3.9. *Let $\mathcal{T}$ be a fixed computably axiomatized true theory extending S^1_2. Then for every $k{\ge}k^{\mathrm{cs}}_{\mathcal{T}}$, one has $S^1_2{+}\phi_{BB}(k){\vdash}Con_{\mathcal{T}}$.*

Proof. Fix $k{\ge}k^{\mathrm{cs}}_{\mathcal{T}}$, and let $U{:=}S^1_2{+}\phi_{BB}(k)$. By the choice of $k^{\mathrm{cs}}_{\mathcal{T}}$, there is a k-state Turing machine computing the same partial function as $E_{\mathcal{T}}$, so we may reason in U about that k-state realization of contradiction search for $\mathcal{T}$.

Since $\phi_{BB}(k)$ asserts the exact value $BB(k){=}t_k$, the theory U proves that every halting k-state Turing machine on blank input halts within at most t_k steps.

Because $\mathcal{T}$ is true, the actual computation of $E_{\mathcal{T}}$ does not find a contradiction within the first t_k steps. This is a fixed finite computation, so by standard bounded-arithmetic formalization of finite computations, S^1_2 proves the corresponding bounded halting fact, and hence so does U. Therefore $U{\vdash}\neg\exists s{\le}t_k\,\big(E_{\mathcal{T}} \text{ halts in exactly } s \text{ steps}\big)$.

Combining these two facts, U proves that the contradiction-search machine for $\mathcal{T}$ never halts at all, that is, $U{\vdash}\neg\exists s\,\big(E_{\mathcal{T}} \text{ halts in exactly } s \text{ steps}\big)$. By the definition of $E_{\mathcal{T}}$, this is exactly $Con_{\mathcal{T}}$. Therefore $S^1_2{+}\phi_{BB}(k){\vdash}Con_{\mathcal{T}}$. □

In particular, if $\mathcal{T}$ is consistent, then for every $k{\ge}k^{\mathrm{cs}}_{\mathcal{T}}$, the true Busy Beaver sentence $\phi_{BB}(k)$ is unprovable in $\mathcal{T}$, since otherwise $\mathcal{T}$ would prove $Con_{\mathcal{T}}$.

With this result, we can show that the existence of some hard true extension $\mathcal{S}{+}\phi$ is equivalent to eventual hardness throughout the Busy Beaver family:

Theorem 3.10. *The following are equivalent:*

1. *There exists a true sentence ϕ such that for every constant c, $\mathcal{S} \overset{n^c}{\nvdash} Con_{\mathcal{S}+\phi}(n)$.*
2. *For all sufficiently large k and every constant c, $\mathcal{S} \overset{n^c}{\nvdash} Con_{S^1_2+\phi_{BB}(k)}(n)$.*

Proof. (1)→(2) Assume there exists a true sentence ϕ such that for every constant c, $\mathcal{S} \overset{n^c}{\nvdash} Con_{\mathcal{S}+\phi}(n)$. Apply Lemma 3.9 with $\mathcal{T}{=}\mathcal{S}{+}\phi$. Fix sufficiently large k, and write $U_k := S^1_2{+}\phi_{BB}(k)$. Then $U_k{\vdash}Con_{\mathcal{S}+\phi}$.

Fix once and for all a proof π_k of $Con_{\mathcal{S}+\phi}$ in U_k. There is a uniform proof transformation sending any $(\mathcal{S}{+}\phi)$-proof of contradiction to a U_k-proof of contradiction: given a code of an $(\mathcal{S}{+}\phi)$-proof of $0{=}1$ of length at most n, one appends the fixed derivation π_k of $Con_{\mathcal{S}+\phi}$ and the standard verification that the coded object is such a proof. This increases proof length by at most

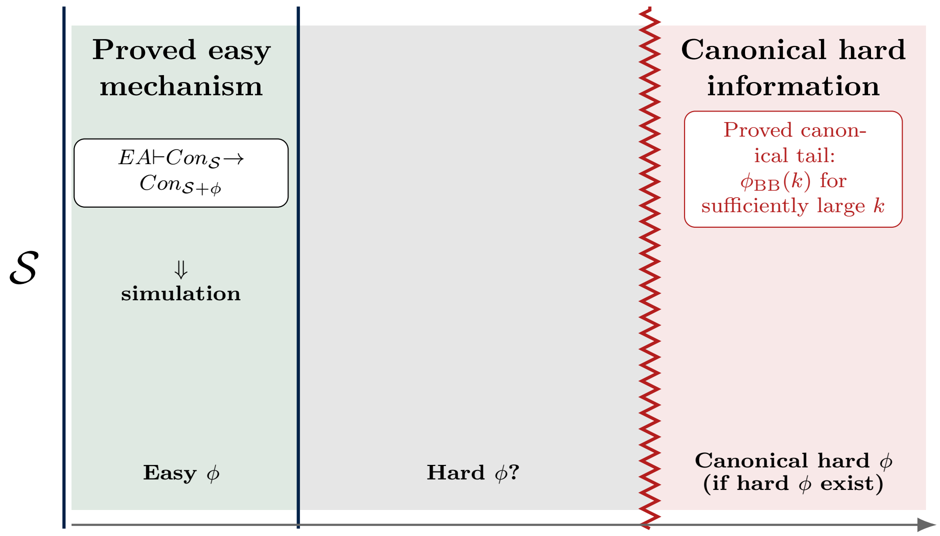

Figure 2: The figure updates Figure 1 to give an equivalent sufficient condition for easy ϕ (Theorem 3.4) and to show that Busy Beavers are a canonical hard-information frontier (Theorem 3.10).

a linear factor. Hence there exist a linear polynomial p_k and a constant d_k such that $\mathcal{S}\vdash^{n^{d_k}} Con_{U_k}(p_k(n)) \to Con_{\mathcal{S}+\phi}(n)$.

Suppose toward a contradiction that $\mathcal{S}\vdash^{n^{O(1)}} Con_{U_k}(n)$ for some sufficiently large k. Since p_k is linear, substituting $p_k(n)$ for n still gives $\mathcal{S}\vdash^{n^{O(1)}} Con_{U_k}(p_k(n))$. Composing these proofs with this implication yields $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$, contrary to the hypothesis.

Therefore for all sufficiently large k and every constant c, $\mathcal{S} \nvdash^{n^c} Con_{S^1_2+\phi_{BB}(k)}(n)$.

(2)$\to$(1) The sentence $\phi_{BB}(k)$ is true for every k, so take $\phi := \phi_{BB}(k)$ for any sufficiently large k supplied by (2). Since every axiom of S^1_2 is an axiom of $\mathcal{S}$, any proof from $S^1_2+\phi$ is verbatim a proof from $\mathcal{S}+\phi$, and $\mathcal{S}$ verifies this inclusion with polynomial overhead; hence polynomial-size $\mathcal{S}$-proofs of $Con_{\mathcal{S}+\phi}(n)$ would yield polynomial-size $\mathcal{S}$-proofs of $Con_{S^1_2+\phi}(n)$, contradicting (2). □

Figure 2 updates Figure 1 to reflect Theorems 3.4 and 3.10. Theorem 3.10 yields the following consequence. If a fixed sound theory $\mathcal{S}$ fails to simulate some true extension $\mathcal{S}+\phi$, then for all sufficiently large k the fixed Busy

Beaver family $(Con_{S^1_2+\phi_{BB}(k)}(n))_n$ is already hard for $\mathcal{S}$. By itself this gives, for each $\mathcal{S}$, only a theory-dependent diagonal family. Under the global hypothesis that every sound theory in the class has some hard true extension, however, one can diagonalize simultaneously over all theory/exponent pairs and obtain a single nonconstructively chosen function $k(n)$ such that the family $(Con_{S^1_2+\phi_{BB}(k(n))}(n))_n$ is hard for every theory in the class. Via the usual passage from bounded-consistency hardness to propositional hardness, this yields a family of tautologies hard for every proof system arising from such a theory class.

Theorem 3.11. *Assume that for every sound theory $\mathcal{S}\supseteq S^1_2$ with polynomial-time decidable axioms there exists a true sentence ψ such that for every constant c, $\mathcal{S} \not\vdash^{n^c} Con_{\mathcal{S}+\psi}(n)$. Then there exists a nonconstructively chosen function $k{:}\mathbb{N}\to\mathbb{N}$ with unbounded range such that for every sound theory $\mathcal{S}\supseteq S^1_2$ with polynomial-time decidable axioms, the family $(Con_{S^1_2+\phi_{BB}(k(n))}(n))_n$ is hard for $\mathcal{S}$. Equivalently, for every such $\mathcal{S}$ and every constant c, $\mathcal{S} \not\vdash^{n^c} Con_{S^1_2+\phi_{BB}(k(n))}(n)$.*

Proof. Let $(\mathcal{S}_e, d_e)_{e\in\mathbb{N}}$ be an enumeration of all pairs consisting of a sound theory $\mathcal{S}_e\supseteq S^1_2$ with polynomial-time decidable axioms and a positive integer exponent $d_e\geq 1$, with each pair appearing infinitely often.

For each e, by hypothesis there exists a true sentence ψ_e such that for every constant c, $\mathcal{S}_e \not\vdash^{n^c} Con_{\mathcal{S}_e+\psi_e}(n)$. Therefore, by Theorem 3.10, there exists K_e such that for every fixed $m\geq K_e$, $\mathcal{S}_e \not\vdash^{\mathcal{O}(n^{d_e})} Con_{S^1_2+\phi_{BB}(m)}(n)$.

We choose inductively sequences $m_0<m_1<m_2<\cdots$ and $n_0<n_1<n_2<\cdots$. At stage $e\geq 1$, choose m_e with $m_e\geq K_e$ and $m_e>m_{e-1}$. Since $\mathcal{S}_e \not\vdash^{\mathcal{O}(n^{d_e})} Con_{S^1_2+\phi_{BB}(m_e)}(n)$, there exists n_e such that every $\mathcal{S}_e$-proof of $Con_{S^1_2+\phi_{BB}(m_e)}(n_e)$ has size greater than $e n_e^{d_e}$. Increase n_e if necessary so that $n_e>n_{e-1}$.

Define $k(n_e) := m_e$ for each $e\geq 1$, and define $k(n)$ arbitrarily on all other inputs. Since (m_e) is strictly increasing, the range of k is unbounded.

Fix a sound theory $\mathcal{S}\supseteq S^1_2$ with polynomial-time decidable axioms, and suppose toward a contradiction that for some constant c, $\mathcal{S}\vdash^{n^c} Con_{S^1_2+\phi_{BB}(k(n))}(n)$. Then there exist constants A, N such that for all $n\geq N$, there is an $\mathcal{S}$-proof of $Con_{S^1_2+\phi_{BB}(k(n))}(n)$ of size at most An^c.

Choose e such that $\mathcal{S}_e=\mathcal{S}$, $d_e\geq c$, $e\geq A$, and $n_e\geq N$; this is possible because every theory appears infinitely often paired with arbitrarily large integers. At $n=n_e$ we have $k(n_e)=m_e$, so $Con_{S^1_2+\phi_{BB}(k(n_e))}(n_e) = Con_{S^1_2+\phi_{BB}(m_e)}(n_e)$. By construction every $\mathcal{S}$-proof of this sentence has size greater than $en_e^{d_e}$, while $en_e^{d_e} \geq An_e^c$. This contradicts the assumed upper bound.

Therefore, for every sound theory $\mathcal{S} \supseteq S^1_2$ with polynomial-time decidable axioms and every constant c, $\mathcal{S} \not\vdash^{n^c} Con_{S^1_2+\phi_{BB}(k(n))}(n)$. $\square$

Under Pudlák's Conjecture, one can give a stronger statement indicating just how large k must be:

Theorem 3.12. *Assume Pudlák's Conjecture. Then for every $k \geq k^{\mathrm{cs}}_{\mathcal{S}}$, for every constant c, $\mathcal{S} \not\vdash^{n^c} Con_{\mathcal{S}+\phi_{BB}(k)}(n)$.*

Proof. Fix $k \geq k^{\mathrm{cs}}_{\mathcal{S}}$. Since $\mathcal{S}+\phi_{BB}(k)$ extends $S^1_2+\phi_{BB}(k)$, Lemma 3.9 implies $\mathcal{S}+\phi_{BB}(k) \vdash Con_{\mathcal{S}}$.

Fix once and for all a proof π_k of $Con_{\mathcal{S}}$ in $\mathcal{S}+\phi_{BB}(k)$. Since k is fixed, the size of π_k is a constant independent of n.

There is a uniform proof transformation sending any $(\mathcal{S}+Con_{\mathcal{S}})$-proof to an $(\mathcal{S}+\phi_{BB}(k))$-proof by replacing each use of the extra axiom $Con_{\mathcal{S}}$ by the fixed derivation π_k. This increases proof length by at most a constant multiplicative and additive factor. Hence there exist a linear polynomial p_k and a constant d_k such that $\mathcal{S} \vdash^{n^{d_k}} Con_{\mathcal{S}+\phi_{BB}(k)}(p_k(n)) \rightarrow Con_{\mathcal{S}+Con_{\mathcal{S}}}(n)$.

Suppose toward a contradiction that $\mathcal{S} \vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi_{BB}(k)}(n)$. Since p_k is linear, substituting $p_k(n)$ for n still gives $\mathcal{S} \vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi_{BB}(k)}(p_k(n))$. Composing these proofs with the displayed implication yields $\mathcal{S} \vdash^{n^{O(1)}} Con_{\mathcal{S}+Con_{\mathcal{S}}}(n)$, contrary to the hypothesis.

Therefore $\mathcal{S} \not\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi_{BB}(k)}(n)$. Since $k \geq k^{\mathrm{cs}}_{\mathcal{S}}$ was arbitrary, this holds for every such k. $\square$

These results show that Busy Beaver hardness is not merely a convenient encoding of difficult extensions, but reflects a structural barrier to simulation. In each case, non-simulation arises because $\mathcal{S}+\phi$ carries consistency-strength information whose transfer is not already visible over weak arithmetic.

Informally, any argument showing that a sound theory $\mathcal{S}$ fails to simulate some true extension $\mathcal{S}+\phi$ also yields, for that same $\mathcal{S}$, some Busy Beaver sentence $\phi_{BB}(k)$ that $\mathcal{S}$ cannot prove witnessing the same obstruction. In this sense, the source of non-simulation already appears among canonical incompleteness statements, and reflects a limit on the ability of $\mathcal{S}$ to exploit true information it cannot itself prove. This is exactly the pattern predicted by **Higher Relative Consistency**: if $EA \nvdash Con_{\mathcal{S}} \rightarrow Con_{\mathcal{S}+\phi}$, then $Con_{\mathcal{S}+\phi}(n)$ should already be hard for $\mathcal{S}$.

3.4 A Busy Beaver Hardness Conjecture

The preceding subsection shows that, if nonsimulation occurs at all, then nonsimulation can be forced into exact Busy Beaver extensions. This is a useful canonicalization result, but it leaves open a sharper question. Can one formulate a Busy Beaver hardness principle that is specific about the parameter k? In other words, instead of saying only that sufficiently large Busy Beaver extensions can witness nonsimulation whenever some hard true extension exists, can we say when the particular extension $\mathcal{S}{+}\phi_{\mathrm{BB}}(k)$ should be hard for $\mathcal{S}$?

The most naive answer is too strong. One might conjecture that, whenever $\mathcal{S}{\nvdash}\phi_{\mathrm{BB}}(k)$, the theory $\mathcal{S}$ cannot feasibly simulate $\mathcal{S}{+}\phi_{\mathrm{BB}}(k)$. The Jeřábek–Pudlák proof-translation mechanism shows that an unprovable true sentence ϕ can nevertheless yield feasible simulation, provided there is a sufficiently strong explanation of the extension by interpretation or relative consistency; in the formulation used here, the relevant weak-base explanation is $EA{\vdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+\phi}$. Thus mere failure of $\mathcal{S}$ to prove $\phi_{\mathrm{BB}}(k)$ is not the right boundary.

The heuristic reason Busy Beaver statements are nevertheless compelling is that exact Busy Beaver values appear to be unusually tight carriers of hidden information. A correct assertion $\phi_{\mathrm{BB}}(k)$ does not merely add an arbitrary true sentence; it constrains all k-state computations. For large k, this information can encode consistency consequences of very strong theories while remaining inaccessible to a weaker base. This suggests that exact Busy Beaver extensions should be hard precisely when their simulation would require using information that the weak base cannot explain.

This leads to the following coarse Busy Beaver special case of the information-constraint viewpoint.

Conjecture 3.13. *(**Busy Beaver Hardness**) Let $\mathcal{S}$ be a sound, finitely axiomatized, sequential theory satisfying the standing assumptions above. Then, for every k, $\mathcal{S}$ does not feasibly simulate $\mathcal{S}{+}\phi_{\mathrm{BB}}(k)$ if and only if $EA{\nvdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}$.*

There is a corresponding Busy-Beaver-restricted version of *Feasible Reflection.* Unlike the coarser tail formulation, this version has no finite-exception threshold built into the statement. It says that, along the exact Busy Beaver family, the weak-base relative-consistency condition is precisely the boundary between feasible simulation and nonsimulation.

Theorem 3.14. *Assume* **Busy Beaver Hardness** *for* $\mathcal{S}$*. Then, for every* k*,* $\mathcal{S}$ *satisfies the following feasible reflection property: if* $\mathcal{S}\overset{n^{O(1)}}{\vdash} Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}(n)$*, then* $EA \vdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}$*. Equivalently, along the exact Busy Beaver family,* $\mathcal{S}\overset{n^{O(1)}}{\vdash} Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}(n)$ *if and only if* $EA \vdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}$*.*

Proof. By **Busy Beaver Hardness**, for every k, $\mathcal{S}$ does not feasibly simulate $\mathcal{S}+\phi_{\mathrm{BB}}(k)$ if and only if $EA \nvdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}$. Taking contrapositives gives that, for every k, $\mathcal{S}$ feasibly simulates $\mathcal{S}+\phi_{\mathrm{BB}}(k)$ if and only if $EA \vdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}$. In the bounded-consistency formulation of feasible simulation, this is exactly the equivalence between $\mathcal{S}\overset{n^{O(1)}}{\vdash} Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}(n)$ and $EA \vdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}$. □

The logical relationship with Pudlák's Conjecture is asymmetric. By Theorem 3.12, Pudlák's Conjecture implies the hard direction of **Busy Beaver Hardness** for every $k \geq k^{\mathrm{cs}}_{\mathcal{S}}$. Indeed, $\mathcal{S}+\phi_{\mathrm{BB}}(k) \vdash Con_{\mathcal{S}}$, so easiness of the Busy Beaver extension would imply easiness of $\mathcal{S}+Con_{\mathcal{S}}$. The converse does not follow: an exact Busy Beaver axiom generally contains substantially more information than $Con_{\mathcal{S}}$, and hardness of the stronger extension need not transfer downward to the weaker one. Thus **Busy Beaver Hardness** is narrower in the family it classifies, while Pudlák's Conjecture supplies a hardness principle that transfers upward to every extension proving $Con_{\mathcal{S}}$. A finer-grained conjecture would be desirable to explain when hardness transfers in the reverse direction.

4 Heuristic Constraints

This section develops several heuristic constraints suggested by the above results. These constraints aim to limit the use of unprovable information in simulations, particularly information drawn from immune sets such as Busy Beavers and Kolmogorov-random strings.

4.1 Tightness

A natural heuristic constraint on criteria for non-simulation is to assert that the best known sufficient condition for simulation is already the best possible. Under this heuristic, the sufficient condition from Theorem 3.4 is also necessary, and therefore becomes a characterization of simulation.

Equivalently, failure of that condition becomes the predicted criterion for non-simulation. We adopt this only as a working principle: the paper does not prove that Theorem 3.4 is optimal, which would in itself resolve open problems.

4.2 Feasible Reflection

A natural structural constraint on simulations is that externally short proofs of bounded consistency should have an internal explanation in a weak base theory. The following definition isolates that requirement.

Definition 4.1. *Feasible Reflection holds for* $\mathcal{S}$ *if for every true sentence* ϕ $\mathcal{S}\overset{n^{O(1)}}{\vdash} Con_{\mathcal{S}+\phi}(n)$ *implies* $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+\phi}$.

Feasible Reflection asserts that external polynomial-size simulation should already admit a relative-consistency explanation over a weak base theory. In light of Theorem 3.2, when $\mathcal{S}\overset{n^{O(1)}}{\vdash} Con_{\mathcal{S}}(n)$, this means that feasible relative consistency should already imply $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+\phi}$. Informally, the point is not that proving each finite instance should automatically yield a proof of the universal statement, but that efficient proofs of all the bounded consistency statements should not exist unless there is already a weak-base explanation for them. In that sense, *Feasible Reflection* is best viewed not as an induction principle, but as a structural constraint on efficient proofs that depend on true but unprovable information.

By Theorem 3.2, if $\mathcal{S}\overset{n^{O(1)}}{\vdash} Con_{\mathcal{S}}(n)$, Feasible Reflection states that feasible relative consistency $(\mathcal{S}\overset{n^{O(1)}}{\vdash} Con_{\mathcal{S}}(p(n))\rightarrow Con_{\mathcal{S}+\phi}(n))$ already yields $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+\phi}$.[8]

4.3 Simulations Employing Unproven Facts

A natural refinement of the Busy Beaver phenomenon is obtained by considering Kolmogorov-random strings. We fix a universal Turing machine U. For each string $x\in\{0,1\}^*$, let $K_U(x)$ denote the plain Kolmogorov complexity

[8] Freund and Pakhomov [4] give an example, based on PA and slow consistency, showing that external efficient simulation does not by itself force interpretability: PA has polynomial-size proofs of $Con_{\mathsf{PA}+\mathrm{Con}^*(\mathsf{PA})}(n)$, while PA does not interpret PA+Con*(PA). This illustrates that additional bridge principles, such as *Feasible Reflection*, are needed to connect simulation with relative-consistency transfer or interpretability.

of x, that is, the length of the shortest program p such that $U(p)=x$. Fix once and for all a constant c_U, and let R denote the set of strings x such that $K_U(x) \geq |x| - c_U$. Thus R is the set of Kolmogorov-random strings relative to U, up to the fixed additive constant c_U. It is well known that R is infinite and immune; any fixed sound effectively axiomatized theory proves $x \in R$ for at most finitely many true instances, by Chaitin's Incompleteness Theorem [3] (see also Li and Vitányi [9]).

Kolmogorov-random axioms $x \in R$ suggest a particularly concrete source of hardness within the relative-consistency viewpoint. The guiding intuition is that efficient simulation should not be able to exploit true random information that the base theory cannot itself recover. This leads to a candidate information-theoretic hardness principle. It should not be viewed as a theorem from the structural discussion above, but rather as a distinguished random-axiom instance of the same general obstruction.

One cannot expect such hardness for every true random axiom. By Theorem 3.4, simulation is possible whenever $EA \vdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+(x\in R)}$. The natural question is therefore whether, in the random-axiom case, the exact obstruction to simulation is failure of this weak-base relative-consistency implication.

Conjecture 4.2. *(*Kolmogorov Hardness, ***KH****) Assume $\mathcal{S}$ is finitely axiomatized and sequential. Whenever $x \in R$ in the standard model, one has $\mathcal{S} \mathrel{\vert\!\!\!\frac{n^{O(1)}}{\quad}} Con_{\mathcal{S}+(x\in R)}(n)$ if and only if $EA \vdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+(x\in R)}$. Equivalently, $\mathcal{S} \mathrel{\vert\!\!\!\frac{n^{O(1)}}{\;/\;}} Con_{\mathcal{S}+(x\in R)}(n)$ if and only if $EA \nvdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+(x\in R)}$.*

Lemma 4.3. *For each fixed string x, EA proves $Con_{\mathcal{S}+(x\in R)} \to x \in R$. Hence, if $EA \vdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+(x\in R)}$, then $EA + Con_{\mathcal{S}} \vdash x \in R$.*

Proof. Fix a string x. The assertion $\neg(x \in R)$ is Σ_1: it is witnessed by a program p and a time t such that $|p| < |x| - c_U$ and $U(p)$ halts in exactly t steps with output x. By provable Σ_1-completeness over EA for the polynomial-time axiomatized theory $\mathcal{S}+(x\in R)$ (Hájek–Pudlák [5]), EA proves that any such witness yields an $\mathcal{S}+(x\in R)$-proof of $\neg(x \in R)$; since $x \in R$ is an axiom of $\mathcal{S}+(x\in R)$, formalized modus ponens turns this into an $\mathcal{S}+(x\in R)$-proof of contradiction. The argument is internal and uniform in the witness, with no case split on whether $x \in R$ holds in the standard model. Thus EA proves $\neg(x \in R) \to \neg Con_{\mathcal{S}+(x\in R)}$, and therefore EA proves $Con_{\mathcal{S}+(x\in R)} \to x \in R$. The consequence follows by composing this implication with $EA \vdash Con_{\mathcal{S}} \to Con_{\mathcal{S}+(x\in R)}$. □

The point of this formulation is that weak-base relative consistency is the exact kind of information that can make simulation possible. Lemma 4.3 shows that provable relative consistency also forces $EA{+}Con_{\mathcal{S}}$ to prove the random axiom. Thus failure of $EA{+}Con_{\mathcal{S}}$ to prove $x{\in}R$ unconditionally rules out the weak-base relative-consistency explanation. This no-access obstruction is incorporated into the master consequence theorem for **Higher Relative Consistency** below, even though the biconditional threshold in Conjecture 4.2 is stated in terms of relative consistency.

Lemma 4.4. *For each fixed string x, if $EA{+}Con_{\mathcal{S}}{\nvdash}x{\in}R$, then $EA{\nvdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+(x\in R)}$.*

Proof. This is the contrapositive of Lemma 4.3. □

5 Proposed Characterization of Simulation

Theorem 3.4 shows that, for $\mathcal{S}$ finitely axiomatized and sequential, relative consistency ($EA{\vdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+\phi}$) implies simulation ($\mathcal{S}{\vdash}^{\!n^{O(1)}}Con_{\mathcal{S}+\phi}(n)$). Taking this theorem to be tight leads to the following structural candidate characterization of simulation. This section develops that proposal and its main consequences. In particular, the case $\phi{=}Con_{\mathcal{S}}$ and the Kolmogorov-randomness axioms arise directly as natural instances of the same general obstruction, while the Busy Beaver reduction shows that arbitrary hard true extensions can be transferred to exact Busy Beaver value statements.

Assumption 5.1. *(**Higher Relative Consistency (HRC)**) Let $\mathcal{S}{\supseteq}S^1_2$ be a sound theory with polynomial-time decidable axioms, and let ϕ be a true sentence. If $EA{\nvdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+\phi}$, then for every constant $c{>}0$, $\mathcal{S}\nvdash^{\!n^c} Con_{\mathcal{S}+\phi}(n)$. Equivalently, efficient simulation of a true extension can occur only when the relative-consistency implication is already provable in EA.*

This assumption is intended as the converse of Theorem 3.4, and hence as a characterization of feasible simulation in this setting. Figure 3 motivates **HRC/FR**: its formal content is not merely that feasible simulation should imply interpretability, but rather that feasible relative consistency and weak-base relative consistency should coincide. The top horizontal equivalence in the figure is the finitely axiomatized Friedman–Visser equivalence. Under the horizontal equivalences in the figure, **HRC/FR** also yields the corresponding left-hand converse from feasible bounded-consistency simulation to interpretability.

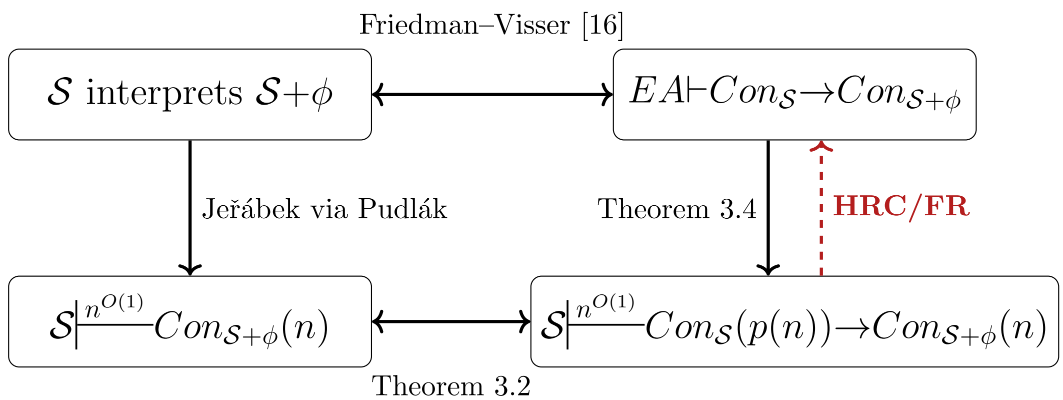

Figure 3: Motivating **HRC/FR**. The dashed arrow is the proposed converse to Theorem 3.4: efficient finite relative-consistency proofs should require weak-base proof of the corresponding relative-consistency implication. Under the horizontal equivalences, this also entails the corresponding left-hand converse from feasible bounded-consistency simulation to interpretability. The figure assumes $\mathcal{S} \supseteq S^1_2$ is sound with polynomial-time decidable axioms, ϕ is true, and $\mathcal{S} \vdash^{n^{O(1)}} Con_{\mathcal{S}}(n)$. For the top equivalence, additionally assume $\mathcal{S}$ is finitely axiomatized and sequential, so that the Friedman–Visser criterion applies.

5.1 Consequences of Higher Relative Consistency

The next theorem summarizes the main properties of **Higher Relative Consistency** and shows that it performs well against the constraints developed earlier. It is tight relative to the best-known simulation result (Theorem 3.4); it is equivalent to *Feasible Reflection*; in the finitely axiomatized sequential case, it implies Pudlák's Conjecture and *Kolmogorov Hardness*; and it packages the no-access random-axiom obstruction from Lemma 4.4.

Theorem 5.2. *Let $\mathcal{S} \supseteq S^1_2$ be a sound theory with polynomial-time decidable axioms such that $\mathcal{S} \vdash^{n^{O(1)}} Con_{\mathcal{S}}(n)$. Then:*

1. *The principles **Higher Relative Consistency** and* Feasible Reflection *are equivalent.*

2. *If, in addition, $\mathcal{S}$ is finitely axiomatized and sequential, then **Higher Relative Consistency** implies Pudlák's Conjecture: for every constant c, $\mathcal{S} \nvdash^{n^c} Con_{\mathcal{S}+Con_{\mathcal{S}}}(n)$.*

3. *If, in addition, $\mathcal{S}$ is finitely axiomatized and sequential, then* ***Higher Relative Consistency*** *is tight relative to Theorem 3.4:* $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$ *if and only if* $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$*.*

4. *If, in addition, $\mathcal{S}$ is finitely axiomatized and sequential, then* ***Higher Relative Consistency*** *implies* Kolmogorov Hardness*.*

5. *Assuming* ***Higher Relative Consistency****, if $x\in R$ is true and* $EA+Con_{\mathcal{S}}\nvdash x\in R$*, then for every constant $c>0$,* $\mathcal{S}\nvdash^{n^c} Con_{\mathcal{S}+(x\in R)}(n)$*.*

Proof. For part (1), first assume **Higher Relative Consistency**. Suppose $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$. If $EA\nvdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$, then **Higher Relative Consistency** would imply $\mathcal{S}\nvdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$, a contradiction. Hence $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$. Thus *Feasible Reflection* holds.

Conversely, assume *Feasible Reflection.* Suppose $EA\nvdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$. If $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$, then by *Feasible Reflection* one would have $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$, a contradiction. Therefore $\mathcal{S}\nvdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$. This is exactly **Higher Relative Consistency**.

For part (2), assume **Higher Relative Consistency**. It suffices to show $EA\nvdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+Con_{\mathcal{S}}}$; **Higher Relative Consistency** then yields that for every constant $c>0$, $\mathcal{S}\nvdash^{n^c} Con_{\mathcal{S}+Con_{\mathcal{S}}}(n)$. Suppose that $\mathcal{S}$ is finitely axiomatized and sequential, and toward a contradiction that $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+Con_{\mathcal{S}}}$. Since $\mathcal{S}+Con_{\mathcal{S}}$ is finitely axiomatized and sequential, the Friedman–Visser interpretability criterion used in the proof of Theorem 3.4, again in the direction supplied by Visser's Interpretation Existence Lemma [16], gives that $\mathcal{S}$ interprets $\mathcal{S}+Con_{\mathcal{S}}$. This contradicts Pudlák's [11] theorem that no consistent sequential theory interprets the theory obtained by adjoining its own consistency statement.

For part (3), assume **Higher Relative Consistency**. If $\mathcal{S}$ is finitely axiomatized and sequential and $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$, then Theorem 3.4 gives $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$.

Conversely, suppose $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$. If $EA\nvdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$, then **Higher Relative Consistency** would imply $\mathcal{S}\nvdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$, a contradiction. Therefore $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+\phi}$.

For part (4), assume **Higher Relative Consistency**, and suppose $\mathcal{S}$ is finitely axiomatized and sequential. Let $x\in R$ be true. By part (3), applied to

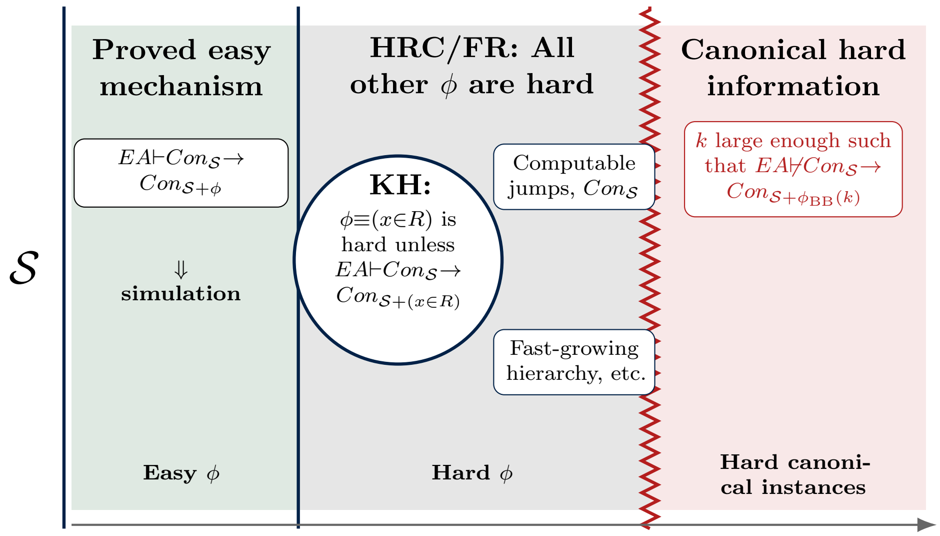

Figure 4: Under **HRC/FR**, complexity theorists have complete information: all ϕ are hard except when $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+\phi}$; $x\in R$ is hard with finite exceptions except when $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+(x\in R)}$ (**KH**); and $\phi_{\mathrm{BB}}(k)$ is hard with finite exceptions except when $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}$.

$\phi\equiv x\in R$, one has $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+(x\in R)}(n)$ if and only if $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+(x\in R)}$. This is exactly *Kolmogorov Hardness* in the form of Conjecture 4.2.

For part (5), assume **Higher Relative Consistency**. By Lemma 4.4, $EA+Con_{\mathcal{S}}\nvdash x\in R$ implies $EA\nvdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+(x\in R)}$. Since $x\in R$ is true, **Higher Relative Consistency** applies with $\phi\equiv x\in R$ and gives that for every constant $c>0$, $\mathcal{S}\nvdash^{n^c} Con_{\mathcal{S}+(x\in R)}(n)$. $\square$

By Chaitin's Incompleteness Theorem applied to the theory $EA+Con_{\mathcal{S}}$, only finitely many true statements of the form $x\in R$ are provable in that theory. Hence for all sufficiently long Kolmogorov-random strings x, the condition $EA+Con_{\mathcal{S}}\nvdash x\in R$ automatically holds, and part (5) gives $\mathcal{S}\nvdash^{n^{O(1)}} Con_{\mathcal{S}+(x\in R)}(n)$.

Figure 4 illustrates that complexity theorists have complete information under **HRC/FR**: the only easy extensions are those whose relative consistency is already visible in the weak base, and every other true extension gives rise to hard bounded-consistency families. More explicitly:

- For an arbitrary true sentence ϕ, the proposed criterion is exact:

$\mathcal{S}$ has feasible proofs of $Con_{\mathcal{S}+\phi}(n)$ only in the explained case $EA{\vdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+\phi}$; otherwise ϕ lies in the hard region.

- For random axioms $\phi{\equiv}(x{\in}R)$, this specialization is **KH**: except for the finite set of weak-base-accessible random facts, $x{\in}R$ yields hard bounded-consistency instances unless $EA{\vdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+(x\in R)}$.

- For exact Busy Beaver facts, the same information constraint gives a canonical hard tail: for sufficiently large k, $\phi_{\mathrm{BB}}(k)$ is hard except in the exceptional cases where $EA{\vdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+\phi_{\mathrm{BB}}(k)}$.

- For Pudlák's conjecture, computable jump operators, and Buss-style jumps, the same picture identifies another source of hard candidates. Khaniki [6] shows that if a computable jump operator exists, then Pudlák's Conjecture holds. A Buss-style jump should be understood as raising absolute consistency strength when it is represented by a true sentence or schema ϕ_{Buss} with $\mathcal{S}{+}\phi_{\mathrm{Buss}}{\vdash}Con_{\mathcal{S}}$. It raises relative-consistency strength, in the sense relevant to **HRC/FR**, only if $EA{\nvdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+\phi_{\mathrm{Buss}}}$. Under that additional relative-consistency failure, **HRC/FR** place ϕ_{Buss} in the hard region.

- Thus **HRC/FR** turn the problem of finding hard tautology families for $\mathcal{S}$ into a relative-consistency classification problem: the easy cases are exactly the weak-base-explained cases, while the unexplained cases produce hard bounded-consistency statements and hence hard propositional translations.

5.2 Robustness of HRC/FR and KH

The formulation of **HRC/FR** is robust under future improvements to the positive simulation theorem. It is conceivable that either the Friedman–Visser interpretability step or the Jeřábek–Pudlák proof-translation step could be strengthened so that more true sentences ϕ become easy. Such a strengthening would not undermine the information-theoretic thesis; it would merely enlarge the class of acceptable explanations. The invariant principle is that *inexplicable efficiency does not exist.* The **KH** consequence is even more stable: any enlarged explanation mechanism still constrained by Chaitin incompleteness can add only finitely many random-axiom exceptions.

Definition 5.3. An *explanation scheme* $\mathfrak{E}$ assigns to each pair $(\mathcal{S}, \phi)$ a condition $\mathsf{Expl}_{\mathfrak{E}}(\mathcal{S}, \phi)$, read as saying that $\mathfrak{E}$ explains why $\mathcal{S}$ should simulate $\mathcal{S}+\phi$. It is *sound for simulation* if $\mathsf{Expl}_{\mathfrak{E}}(\mathcal{S}, \phi)$ implies $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$. It is *Chaitin-bounded for random axioms over* $\mathcal{S}$ if there is a fixed sound effectively axiomatized theory $\mathcal{B}^{\mathcal{S}}_{\mathfrak{E}}\supseteq EA$ such that $\mathsf{Expl}_{\mathfrak{E}}(\mathcal{S}, x\in R)$ implies $\mathcal{B}^{\mathcal{S}}_{\mathfrak{E}}\vdash x\in R$.

The original weak-base relative-consistency explanation is the special case $\mathsf{Expl}_{EA}(\mathcal{S}, \phi)$ iff $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+\phi}$. It is sound for simulation on the class of finitely axiomatized sequential theories by Theorem 3.4. For random axioms it is Chaitin-bounded: if $EA\vdash Con_{\mathcal{S}}\rightarrow Con_{\mathcal{S}+(x\in R)}$, then $EA+Con_{\mathcal{S}}\vdash x\in R$ by Lemma 4.3. Thus one may take $\mathcal{B}^{\mathcal{S}}_{\mathfrak{E}}=EA+Con_{\mathcal{S}}$.

Assumption 5.4. *(**No Inexplicable Efficiency**, relative to $\mathfrak{E}$.) If ϕ is true and $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$, then $\mathsf{Expl}_{\mathfrak{E}}(\mathcal{S}, \phi)$.*

Theorem 5.5. *Let $\mathfrak{E}$ be sound for simulation and assume **No Inexplicable Efficiency** relative to $\mathfrak{E}$. Then, for every true ϕ, $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}+\phi}(n)$ iff $\mathsf{Expl}_{\mathfrak{E}}(\mathcal{S}, \phi)$. If moreover $\mathfrak{E}$ is Chaitin-bounded for random axioms over $\mathcal{S}$, then $\{x\in R : \mathsf{Expl}_{\mathfrak{E}}(\mathcal{S}, x\in R)\}$ is finite, and hence for all but finitely many true $x\in R$, $\mathcal{S}\nvdash^{n^{O(1)}} Con_{\mathcal{S}+(x\in R)}(n)$.*

Proof. The equivalence is immediate: the forward direction is **No Inexplicable Efficiency**, and the reverse direction is soundness of $\mathfrak{E}$ for simulation. For the random-axiom claim, let $\mathcal{B}^{\mathcal{S}}_{\mathfrak{E}}$ witness Chaitin-boundedness. If $\mathsf{Expl}_{\mathfrak{E}}(\mathcal{S}, x\in R)$ holds, then $\mathcal{B}^{\mathcal{S}}_{\mathfrak{E}}\vdash x\in R$. Since $\mathcal{B}^{\mathcal{S}}_{\mathfrak{E}}$ is fixed, sound, and effectively axiomatized, Chaitin incompleteness implies that it proves only finitely many true assertions of the form $x\in R$. Thus only finitely many true random axioms are explained by $\mathfrak{E}$. Outside this finite set, **No Inexplicable Efficiency** gives the stated nonsimulation. $\square$

Thus a stronger positive simulation theorem would merely replace the current weak-base explanation scheme by a larger one. As long as its random-axiom explanations are mediated by some fixed sound effective theory, the cofinite hard tail predicted by **KH** remains unchanged; only the finite exceptional set can grow.

6 Further Conjectural Extensions

The previous section addressed the asymptotic hard families predicted by the no-optimal-proof-system picture, even when one ranges over theories of arbitrarily high consistency strength. This section asks for a strictly stronger finite-scale version: do even the strongest proof systems have ubiquitous small hard tautologies, even if chosen at random? This conjecture can be read as a formalization of the folklore intuition that "for any proof system most formulas are hard," while avoiding the naive proposal that Krajíček describes as "void" for random DNFs [7, Ch. 19]; see also Pitassi [10]. The tautologies here are not sampled formulas. They are indexed by true strings $x{\in}R$, and the conjecture gives a common hardness onset for all such strings of each fixed sufficiently large length.

It is awkward to require hardness at one predetermined small value of n for a specific sentence ϕ, such as $\phi{=}Con_{\mathcal{S}}$ or $\phi_{\mathrm{BB}}(k)$, since a pathological theory $\mathcal{S}$ might contain an axiom asserting $Con_{\mathcal{S}+\phi}(10^{100})$. We therefore fix the length m of a true random axiom $x{\in}R$ and, for each desired exponential saving ϵ, allow a threshold $N^{\mathcal{S}}_{R,m,\epsilon}$ beyond which hardness is required for every bounded-consistency parameter n. The threshold is uniform over all true random axioms of the same length.

A related pairwise question asks whether independently random axioms can help one another. If $x,y{\in}R$ have the same length and each remains random even conditional on the other, should adjoining $y{\in}R$ help prove $Con_{\mathcal{S}+(x\in R)}(n)$? This requires a genuinely pairwise strengthening and is treated after the single-axiom density consequence.

These finite-scale principles are not consequences of **HRC/FR** or **KH**; they are stronger conjectures motivated by that picture. Their purpose is to indicate what additional proof-complexity phenomena would follow if the random-axiom obstruction were strengthened from asymptotic polynomial nonsimulation to exponential hardness beyond a uniform length-dependent onset.

Fix a sound theory $\mathcal{S}$ satisfying the standing assumptions. Let $P^{\mathcal{S}}_R{:=}\{x{\in}R{:}EA{+}Con_{\mathcal{S}}{\vdash}x{\in}R\}$ and let $C^{\mathcal{S}}_{\mathrm{rel}}{:=}\{x{\in}R{:}EA{\vdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+(x\in R)}\}$. By Chaitin's incompleteness theorem, $P^{\mathcal{S}}_R$ is finite. Let $k^{\mathcal{S}}_R$ be any integer exceeding the length of every string in $P^{\mathcal{S}}_R$. By Lemma 4.3, one has $C^{\mathcal{S}}_{\mathrm{rel}}{\subseteq}P^{\mathcal{S}}_R$. Consequently, whenever $m{>}k^{\mathcal{S}}_R$ and $x{\in}R{\cap}\{0,1\}^m$ is true, one has $EA{\nvdash}Con_{\mathcal{S}}{\to}Con_{\mathcal{S}+(x\in R)}$.

For each $m{>}k^{\mathcal{S}}_R$ and each ϵ with $0{<}\epsilon{<}1$, the threshold $N^{\mathcal{S}}_{R,m,\epsilon}$ below may depend on $\mathcal{S}$, the fixed predicate R, the axiom length m, and ϵ, but not on the particular string x or on the later bounded-consistency parameter n. It may absorb the least length of a valid proof string and the syntactic cost of mentioning an axiom $x{\in}R$ of length m.

Conjecture 6.1. ***(SETH-K-Finite)*** *Assume $\mathcal{S}$ is finitely axiomatized and sequential, and assume $\mathcal{S}\vdash^{n^{O(1)}} Con_{\mathcal{S}}(n)$. For every $m>k_R^{\mathcal{S}}$ and every ϵ with $0<\epsilon<1$, there exists $N_{R,m,\epsilon}^{\mathcal{S}}$ such that, for every true $x\in R\cap\{0,1\}^m$ and every $n>N_{R,m,\epsilon}^{\mathcal{S}}$, one has $\mathcal{S}\vdash^{2^{(1-\epsilon)n}} Con_{\mathcal{S}+(x\in R)}(n)$ if and only if $EA\vdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+(x\in R)}$.*

The right-hand side is exactly the weak-base relative-consistency threshold proposed by **HRC/FR**. By the definition of $k_R^{\mathcal{S}}$, it is false throughout the range $m>k_R^{\mathcal{S}}$. Thus, on the quantified tail, **SETH-K-Finite** is equivalently the assertion that, for every $m>k_R^{\mathcal{S}}$ and every ϵ with $0<\epsilon<1$, there exists $N_{R,m,\epsilon}^{\mathcal{S}}$ such that, for every true $x\in R\cap\{0,1\}^m$ and every $n>N_{R,m,\epsilon}^{\mathcal{S}}$, one has $\mathcal{S}\not\vdash^{2^{(1-\epsilon)n}} Con_{\mathcal{S}+(x\in R)}(n)$. We retain the biconditional because it identifies the same information boundary as **HRC/FR**.

The threshold $N_{R,m,\epsilon}^{\mathcal{S}}$ may depend on $\mathcal{S}$, the fixed predicate R, the axiom length m, and the desired exponential saving ϵ, but not on the particular string x or on the later bounded-consistency parameter n. This is the usual **SETH**-style quantifier pattern: for every fixed exponential saving, hardness holds beyond an onset that may depend on that saving.

6.1 Density of Hard Sentences

The first consequence of **SETH-K-Finite** concerns the abundance of hard single-axiom extensions. For density arguments, the fixed-deficiency set R is not the most convenient one, because it need not occupy a fraction tending to 1 inside each length. We therefore apply the same conjectural schema to the logarithmic-deficiency set $R^{\log}$, defined by $x\in R^{\log}$ if and only if $K_U(x)\geq|x|-d\log|x|$. Define $k_{R^{\log}}^{\mathcal{S}}$ and $N_{R^{\log},m,\epsilon}^{\mathcal{S}}$ exactly as above, with $R^{\log}$ in place of R.

Theorem 6.2. *Assume* ***SETH-K-Finite*** *with R replaced by $R^{\log}$. For every $m>k_{R^{\log}}^{\mathcal{S}}$ and every ϵ with $0<\epsilon<1$, there exists $N_{R^{\log},m,\epsilon}^{\mathcal{S}}$ such that, for every $n>N_{R^{\log},m,\epsilon}^{\mathcal{S}}$, at least $(1-O(m^{-d}))2^m$ strings $x\in\{0,1\}^m$ satisfy $x\in R^{\log}$ and $\mathcal{S}\not\vdash^{2^{(1-\epsilon)n}} Con_{\mathcal{S}+(x\in R^{\log})}(n)$.*

Proof. If $x\notin R^{\log}$ and $|x|=m$, then $K_U(x)<m-d\log m$. Hence the number of strings of length m outside $R^{\log}$ is at most $2^{m-d\log m+1}=O(2^m/m^d)$. Therefore at least $(1-O(m^{-d}))2^m$ strings of length m lie in $R^{\log}$. Since $m>k_{R^{\log}}^{\mathcal{S}}$, every true $x\in R^{\log}\cap\{0,1\}^m$ satisfies $EA\nvdash Con_{\mathcal{S}}\to Con_{\mathcal{S}+(x\in R^{\log})}$. For every

$n{>}N^{\mathcal{S}}_{R^{\log},m,\epsilon}$, the conclusion follows from **SETH-K-Finite** with R replaced by $R^{\log}$. □

Thus, for every sufficiently large axiom length m and every ϵ with $0{<}\epsilon{<}1$, once $n{>}N^{\mathcal{S}}_{R^{\log},m,\epsilon}$, hard bounded-consistency instances occupy a $(1{-}O(m^{-d}))$ fraction of $\{0,1\}^m$. If these bounded-consistency statements are translated into tautologies in the usual way, then the resulting tautologies have size $(m{+}n)^{O(1)}$ and are indexed by a density-$1{-}O(m^{-d})$ set of strings of length m.

The density theorem is a direct consequence of the single-axiom conjecture. The next subsection asks the stronger pairwise question of whether one independently random axiom can help prove the bounded consistency of another random-axiom extension.

6.2 Pairwise Hardness and No Mutual Help

Ordinary **SETH-K-Finite** concerns proofs in $\mathcal{S}$ of $Con_{\mathcal{S}+(x\in R)}(n)$. It does not determine whether the different extension $\mathcal{S}{+}(y{\in}R)$ can help prove that same statement. Genuine no mutual help therefore requires a pairwise strengthening.

Fix a universal conditional machine compatible with U and a constant c_U^{cond}. For strings $x,y{\in}\{0,1\}^m$, write $x{\in}R|y$ if $K_U(x|y){\geq}m{-}c_U^{\text{cond}}$. Write $x{\perp_R}y$ if $x,y{\in}R$, $x{\in}R|y$, and $y{\in}R|x$. Thus $x{\perp_R}y$ says that each string remains random when the other is supplied as auxiliary information.

The following theorem identifies the information that a weak-base relative-consistency implication would provide.

Theorem 6.3. *For every pair of strings x,y, if $EA{\vdash}Con_{\mathcal{S}+(y\in R)}{\to}Con_{\mathcal{S}+(x\in R)}$, then $EA{+}Con_{\mathcal{S}+(y\in R)}{\vdash}x{\in}R$.*

Proof. By Lemma 4.3, one has $EA{\vdash}Con_{\mathcal{S}+(x\in R)}{\to}x{\in}R$. Composing this implication with $EA{\vdash}Con_{\mathcal{S}+(y\in R)}{\to}Con_{\mathcal{S}+(x\in R)}$ gives $EA{\vdash}Con_{\mathcal{S}+(y\in R)}{\to}x{\in}R$, which is equivalent to $EA{+}Con_{\mathcal{S}+(y\in R)}{\vdash}x{\in}R$. □

Corollary 6.4. *For every pair of strings x,y, if $EA{+}Con_{\mathcal{S}+(y\in R)}{\nvdash}x{\in}R$, then $EA{\nvdash}Con_{\mathcal{S}+(y\in R)}\ {\to}Con_{\mathcal{S}+(x\in R)}$.*

Proof. This is the contrapositive of Theorem 6.3. □

The remaining step is not supplied by conditional Kolmogorov randomness alone. Although $x{\in}R|y$ says that the bits of y do not provide a short description of x, the sentence $Con_{\mathcal{S}+(y\in R)}$ might in principle carry information not reducible to the literal bits of y. The required factual nonaccess principle is therefore stated explicitly.

Conjecture 6.5. *(**Conditional Factual No Access**) Fix a sound theory $\mathcal{S}$ satisfying the standing assumptions. There exists $k^{\mathcal{S}}_{R,\mathrm{access}}$ such that, for every $m>k^{\mathcal{S}}_{R,\mathrm{access}}$ and every $x,y\in\{0,1\}^m$ satisfying $x\perp_R y$, one has $EA+Con_{\mathcal{S}+(y\in R)}\nvdash x\in R$ and $EA+Con_{\mathcal{S}+(x\in R)}\nvdash y\in R$.*

Thus the conjectural content is that, even after the weak base is given the consistency of the extension containing $y\in R$, it still lacks access to the independent fact $x\in R$, and conversely. The desired relative-consistency obstruction follows formally.

Theorem 6.6. *Assume **Conditional Factual No Access**. Then there exists $k^{\mathcal{S}}_{R,\mathrm{access}}$ such that, for every $m>k^{\mathcal{S}}_{R,\mathrm{access}}$ and every $x,y\in\{0,1\}^m$ satisfying $x\perp_R y$, one has $EA\nvdash Con_{\mathcal{S}+(y\in R)}\rightarrow Con_{\mathcal{S}+(x\in R)}$ and $EA\nvdash Con_{\mathcal{S}+(x\in R)}\rightarrow Con_{\mathcal{S}+(y\in R)}$.*

Proof. Fix $m>k^{\mathcal{S}}_{R,\mathrm{access}}$ and $x,y\in\{0,1\}^m$ satisfying $x\perp_R y$. By **Conditional Factual No Access**, $EA+Con_{\mathcal{S}+(y\in R)}\nvdash x\in R$. Corollary 6.4 therefore gives $EA\nvdash Con_{\mathcal{S}+(y\in R)}\rightarrow Con_{\mathcal{S}+(x\in R)}$. Interchanging x and y gives the reverse non-implication. $\square$

Theorem 6.6 supplies the information-theoretic obstruction. A separate finite-hardness principle is needed to convert failure of weak-base relative consistency into exponential proof lower bounds. The following is the pairwise analog of **SETH-K-Finite**.

Conjecture 6.7. *(**Pairwise SETH-K-Finite**) Fix a sound, finitely axiomatized, sequential theory $\mathcal{S}$ satisfying the standing assumptions. There exists $k^{\mathcal{S}}_{R,\mathrm{hard}}$ such that, for every $m>k^{\mathcal{S}}_{R,\mathrm{hard}}$ and every ϵ with $0<\epsilon<1$, there exists $N^{\mathcal{S},\mathrm{pair}}_{R,m,\epsilon}$ such that, for every $x,y\in\{0,1\}^m$ satisfying $x\perp_R y$ and every $n>N^{\mathcal{S},\mathrm{pair}}_{R,m,\epsilon}$, one has $\mathcal{S}+(y\in R)\vdash^{\!2^{(1-\epsilon)n}} Con_{\mathcal{S}+(x\in R)}(n)$ if and only if $EA\vdash Con_{\mathcal{S}+(y\in R)}\rightarrow Con_{\mathcal{S}+(x\in R)}$.*

Because $x\perp_R y$ is symmetric and the threshold depends only on m, the same conjecture applied to the ordered pair (y,x) gives the reverse comparison.

Theorem 6.8. *Assume **Conditional Factual No Access** and **Pairwise SETH-K-Finite**, and let $k^{\mathcal{S}}_{R,\mathrm{pair}}:=\max\{k^{\mathcal{S}}_{R,\mathrm{access}},k^{\mathcal{S}}_{R,\mathrm{hard}}\}$. Then, for every*

$m{>}k^{\mathcal{S}}_{R,\mathrm{pair}}$ and every ϵ with $0{<}\epsilon{<}1$, there exists $N^{\mathcal{S},\mathrm{pair}}_{R,m,\epsilon}$ such that, for every $x,y{\in}\{0,1\}^m$ satisfying $x{\perp_R}y$ and every $n{>}N^{\mathcal{S},\mathrm{pair}}_{R,m,\epsilon}$, one has $\mathcal{S}{+}(y{\in}R) \not\vdash^{2^{(1-\epsilon)n}} Con_{\mathcal{S}+(x\in R)}(n)$ and $\mathcal{S}{+}(x{\in}R) \not\vdash^{2^{(1-\epsilon)n}} Con_{\mathcal{S}+(y\in R)}(n)$.

Proof. Fix $m{>}k^{\mathcal{S}}_{R,\mathrm{pair}}$ and $x,y{\in}\{0,1\}^m$ satisfying $x{\perp_R}y$. By Theorem 6.6, $EA{\nvdash}Con_{\mathcal{S}+(y\in R)} \to Con_{\mathcal{S}+(x\in R)}$. The biconditional in **Pairwise SETH-K-Finite** therefore gives $\mathcal{S}{+}(y{\in}R) \not\vdash^{2^{(1-\epsilon)n}} Con_{\mathcal{S}+(x\in R)}(n)$ for every $0{<}\epsilon{<}1$ and every $n{>}N^{\mathcal{S},\mathrm{pair}}_{R,m,\epsilon}$. Interchanging x and y gives the reverse lower bound. □

The conclusion concerns the original target $Con_{\mathcal{S}+(x\in R)}(n)$: adjoining $y{\in}R$ does not give short proofs of that bounded-consistency statement. This is stronger and more specific than saying only that the joint extension $\mathcal{S}{+}(x{\in}R){+}(y{\in}R)$ remains hard.

6.3 Density of No-Mutual-Help Pairs

The pairwise conjectures also have a density consequence when randomness is measured with logarithmic deficiency. For equal-length strings $x,y{\in}\{0,1\}^m$, write $x{\perp_{R^{\log}}}y$ if $x,y{\in}R^{\log}$, $K_U(x|y){\ge}m{-}d\log m$, and $K_U(y|x){\ge}m{-}d\log m$.

Lemma 6.9. *For every sufficiently large m, at least $(1{-}O(m^{-d}))2^{2m}$ ordered pairs $(x,y){\in}\{0,1\}^m{\times}\{0,1\}^m$ satisfy $x{\perp_{R^{\log}}}y$.*

Proof. For each fixed $y{\in}\{0,1\}^m$, fewer than $2^{m-d\log m}$ strings x satisfy $K_U(x|y){<}m{-}d\log m$. Hence the number of ordered pairs failing $K_U(x|y){\ge}m{-}d\log m$ is $O(2^{2m}/m^d)$. The same estimate applies with x and y interchanged. The ordinary counting bound gives the same estimate for pairs in which $x{\notin}R^{\log}$ or $y{\notin}R^{\log}$. A union bound gives the conclusion. □

Corollary 6.10. *Assume* ***Conditional Factual No Access*** *and* ***Pairwise SETH-K-Finite*** *with R replaced by $R^{\log}$, and let $k^{\mathcal{S}}_{R^{\log},\mathrm{pair}}{:=}\max\{k^{\mathcal{S}}_{R^{\log},\mathrm{access}},k^{\mathcal{S}}_{R^{\log},\mathrm{hard}}\}$. Then, for every sufficiently large $m{>}k^{\mathcal{S}}_{R^{\log},\mathrm{pair}}$ and every ϵ with $0{<}\epsilon{<}1$, there exists $N^{\mathcal{S},\mathrm{pair}}_{R^{\log},m,\epsilon}$ such that, for every $n{>}N^{\mathcal{S},\mathrm{pair}}_{R^{\log},m,\epsilon}$, at least $(1{-}O(m^{-d}))2^{2m}$ ordered pairs $(x,y){\in}\{0,1\}^m{\times}\{0,1\}^m$ satisfy $\mathcal{S}{+}(y{\in}R^{\log}) \not\vdash^{2^{(1-\epsilon)n}} Con_{\mathcal{S}+(x\in R^{\log})}(n)$ and $\mathcal{S}{+}(x{\in}R^{\log}) \not\vdash^{2^{(1-\epsilon)n}} Con_{\mathcal{S}+(y\in R^{\log})}(n)$.*

Proof. By Lemma 6.9, a $(1{-}O(m^{-d}))$ fraction of ordered pairs of length m are mutually conditionally random with respect to $R^{\log}$. Apply Theorem 6.8 with $R^{\log}$ in place of R. □

7 Remarks on Provability

A complementary question is whether **Higher Relative Consistency** might itself be unprovable. This section is exploratory: it does not establish an independence result for **Higher Relative Consistency**, but records reasons one might expect that natural uniform formulations of the principle will be difficult to prove inside weak first-order theories. The first reason is model-theoretic: the intended meaning of **Higher Relative Consistency** is external and specific to the standard model, whereas any proof in weak arithmetic must proceed through the internal arithmetized proof predicate. The second is complexity-theoretic: a sufficiently uniform proof of **Higher Relative Consistency** would already yield the major lower bounds that motivate the paper.

The model-theoretic issue is basic. Let $Prf_{\mathcal{S}}(p,q)$ be the usual formula expressing that p codes an $\mathcal{S}$-proof of the sentence with Gödel number q. If $M{\models}\mathcal{S}$ is nonstandard, then an element $p{\in}M$ may satisfy $M{\models}Prf_{\mathcal{S}}(p,\ulcorner\varphi\urcorner)$ even though p is nonstandard and does not correspond to any genuine finite proof in the standard model. From the internal point of view of M, however, such a p is simply a proof of φ. Thus nonstandard models can contain spurious witnesses to provability that have no external proof-theoretic meaning.

This matters because the assertion that $\mathcal{S}\overset{n^c}{\vdash} Con_{\mathcal{S}+\phi}(n)$ is intended externally. It means that for each standard n there is a genuine finite $\mathcal{S}$-proof of the standard sentence $Con_{\mathcal{S}+\phi}(n)$ whose length is bounded by a fixed polynomial in n. Any internal first-order formalization, however, necessarily replaces this by quantification over all elements of a model, including nonstandard proof codes and nonstandard bounds. Thus the internalized sentence can hold in a nonstandard model for reasons that have nothing to do with genuine external simulation.

Accordingly, **Higher Relative Consistency** is not merely a statement about the absence of internal witnesses to short proofs. It is a statement that internal proof data correctly tracks external simulation data in the standard model. But that is precisely what first-order arithmetic cannot in general enforce from within: it has no way to isolate the standard proof codes from nonstandard ones, nor the standard proof-length bounds from nonstandard ones. In this sense, **Higher Relative Consistency** behaves less like an ordinary reflection principle and more like a standardness principle saying that only genuine external polynomial-size proofs should count as simulation data.

The same issue appears in the random-axiom setting. Even when $\mathcal{S}\nvdash x{\in}R$, a nonstandard model may still internally treat the bounded-consistency statements for $\mathcal{S}{+}(x{\in}R)$ as having short $\mathcal{S}$-proofs, because nonstandard proof codes can witness the corresponding arithmetized proof predicates. Thus an internalized version of the simulation claim may hold in such a model even when no genuine polynomial simulation exists externally. This is exactly why principles such as **Higher Relative Consistency** and *Kolmogorov Hardness* are best understood as external constraints on when apparent simulation data should be taken seriously.

There is also a second reason to doubt provability. A sufficiently uniform proof of **Higher Relative Consistency** would not be a modest metamathematical tidying-up result. Combined with the unconditional simulation theorem and the Busy Beaver reduction, it would already yield sweeping lower bounds, including canonical Busy Beaver witnesses to non-simulation and, in the global form, the nonexistence of an optimal theory or proof system. Thus any proof of **Higher Relative Consistency** strong enough to support the paper's main applications would already settle central open problems.

This point is especially clear in the uniform setting. For a fixed pair $(\mathcal{S}, \phi)$, an instance of **Higher Relative Consistency** can be arithmetized as an ordinary first-order sentence about proof predicates and proof lengths. Thus, for a specific theory such as ZFC, it is meaningful to ask whether that particular instance is provable, refutable, or independent. However, the discussion above does not establish such an independence result for any fixed instance. Rather, it explains why the most natural and useful forms of **Higher Relative Consistency** are not merely ordinary internal first-order assertions.

The difficulty increases once one passes from fixed instances to uniform forms. A theorem ranging over all true ϕ, even for one fixed $\mathcal{S}$, must still relate internal proof-predicate assertions to genuine standard-model simulation behavior. That is exactly where nonstandard models interfere. What is difficult is not writing down the arithmetization, but justifying the passage from internal proof data to external polynomial simulation claims. The difficulty becomes sharper still in the fully global form ranging over all sound $\mathcal{S}$ and all true ϕ, since both soundness and truth are themselves external semantic conditions.

This also clarifies the role of stronger foundational frameworks. Merely moving from first-order to second-order syntax does not by itself remove the problem, so long as one remains inside a formal theory with its own internal proof predicate and associated nonstandard proof theory. The real issue is not the order of the language, but access to standard truth. For exact Busy Beaver sentences $\phi_{BB}(k)$, the obstruction is already visible: a uniform treatment of all such sentences requires more than the resources needed merely to formalize bounded consistency. The upper-bound side of an exact Busy Beaver statement is naturally Π_1, while the lower-bound side is witnessed by an explicit finite halting computation. Thus a metatheory with a truth predicate for all true Π_1 sentences, together with ordinary arithmetic verification of finite computations, is a natural setting for uniform reasoning about the Busy Beaver family. For **Higher Relative Consistency** in full generality, ranging over arbitrary true ϕ, one would want a broader truth or satisfaction framework, or else an explicitly external axiom schema.

Accordingly, the right conclusion is not that the preceding argument proves independence of **Higher Relative Consistency** from ZFC or from any other specific theory. Rather, it shows that the global form of **Higher Relative Consistency** is most naturally understood as an external structural principle whose faithful formulation already points beyond ordinary first-order arithmetic. If **Higher Relative Consistency** is true, its role may therefore be closer to that of an external axiom governing when apparent simulation data reflects genuine proof-theoretic structure.

8 Conclusion

This paper studies the Characterization Problem for simulation between arithmetic theories and advances a specific conjectural answer to its hardness side. It does not solve the problem. Rather, its unconditional results isolate constraints that any successful characterization should satisfy, while its conjectural part proposes one way those constraints might fit together.

The paper's two unconditional contributions are structural. First, feasible relative consistency suffices for simulation, giving a general upper-bound mechanism that includes the case in which $EA \vdash Con_{\mathcal{S}} \rightarrow Con_{\mathcal{S}+\phi}$. Second, any hard true extension can be replaced by one of Busy Beaver form, showing that canonical incompleteness phenomena already suffice to witness non-simulation.

Against this backdrop, the paper proposes **Higher Relative Consistency** as its central structural criterion: if $EA \nvdash Con_{\mathcal{S}} \rightarrow Con_{\mathcal{S}+\phi}$, then for every constant $c>0$, $\mathcal{S} \not\vdash^{n^c} Con_{\mathcal{S}+\phi}(n)$. This is the paper's main candidate characterization of simulation in terms of relative-consistency transfer in a weak base theory.

Within this framework, the paper also formulates *Kolmogorov Hardness* as a distinguished random-axiom instance of the same general obstruction. Its guiding idea is that canonical random axioms should instantiate the hardness predicted by failure of weak-base relative-consistency transfer. In this sense, *Kolmogorov Hardness* is not a competing criterion, but a particularly natural specialization of **Higher Relative Consistency**.

Proving **Higher Relative Consistency** would amount to a major breakthrough. As the missing converse to the strongest known upper-bound mechanism, it would convert relative-consistency transfer in a weak base theory from a sufficient condition into a genuine characterization of simulation. Whether that proposal is correct remains open, but the present paper substantially narrows the space of plausible alternatives: any successful characterization must explain sufficient conditions for simulation and why canonical incompleteness phenomena, already at the level of Busy Beaver truths and random axioms, suffice to witness its failure.

References

[1] Scott Aaronson, *The busy beaver frontier*, SIGACT News **51** (2020), no. 3, 32–54.

[2] Samuel R. Buss, *Bounded arithmetic*, Lecture notes, Bibliopolis, 1986.

[3] Gregory J. Chaitin, *Information-theoretic limitations of formal systems*, JACM **21** (1974), no. 3, 403–424.

[4] Anton Freund and Fedor Pakhomov, *Short proofs for slow consistency*, Notre Dame Journal of Formal Logic **61** (2020), no. 1, 31–49.

[5] Petr Hájek and Pavel Pudlák, *Metamathematics of first-order arithmetic*, Perspectives in Logic, Cambridge University Press, 1998.

[6] Erfan Khaniki, *Jump operators, interactive proofs and proof complexity generators*, 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS), 2024, pp. 573–593.

[7] Jan Krajíček, *Proof complexity*, Cambridge University Press, New York, NY, 2019.

[8] Jan Krajíček and Pavel Pudlák, *Propositional proof systems, the consistency of first order theories and the complexity of computations*, The Journal of Symbolic Logic **54** (1989), no. 3, 1063–1079.

[9] Ming Li and Paul M. B. Vitányi, *An introduction to Kolmogorov complexity and its applications*, Texts in Computer Science, Springer, 2008.

[10] Toni Pitassi, *Proof complexity and meta-complexity tutorial(2)*, `https://www.youtube.com/watch?v=-9hwU1HtfHM&t=1135s`.

[11] Pavel Pudlák, *Cuts, consistency statements and interpretations*, Journal of Symbolic Logic **50** (1985), no. 2, 423–441.

[12] ______, *On the length of proofs of finitistic consistency statements in first order theories*, Studies in Logic and the Foundations of Mathematics, vol. 120, Elsevier, 1986, pp. 165–196.

[13] ______, *Logical foundations of mathematics and computational complexity: A gentle introduction*, Springer, 2013.

[14] ______, *Incompleteness in the finite domain*, Bull. Symb. Log. **23** (2017), no. 4, 405–441.

[15] Tibor Rado, *On non-computable functions*, The Bell System Technical Journal **41** (1962), no. 3, 877–884.

[16] Albert Visser, *The interpretation existence lemma*, Feferman on Foundations: Logic, Mathematics, Philosophy (Gerhard Jäger and Wilfried Sieg, eds.), Springer International Publishing, Cham, 2017, pp. 101–144.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il