EFFICIENT PROVER–ESTIMATOR DEBATE VIA SIGNED LEVEL-SET OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Recursive debate makes complex claims verifiable by reducing them to subclaims that a human can eventually judge. Its weakness is obfuscation: a dishonest prover can hide an error among subclaims that an honest opponent cannot efficiently identify. Brown-Cohen et al. (2026) introduced *prover–estimator* debate, replacing the search for a false subclaim with probabilistic predictions over all submitted subclaims. Their work showed that suitable probability estimators exist, but their argument uses reference answers unavailable to the estimator, which they note makes it non-executable, and provides a circuit-size upper bound exponential in the branching width. We here explore the conditions under which estimators can instead be constructed efficiently. When decompositions cannot profitably change the meaning of a claim, estimator construction reduces to local optimization guided by whether the prover alleges that the current estimate is too high or too low. Assuming efficient signed level-set optimization, our algorithm uses resettable access to a fixed prover and simulated debate scores, never hidden reference answers, and outputs an estimator satisfying a soundness bound against that prover, with construction and evaluation costs polynomial in the mechanism's parameters. Conversely, without this semantic restriction, a uniform compiler for a promised class of instances would yield randomized polynomial-time algorithms for every polynomial-time verifiable total search relation, even when the local gates are projections or constants. Thus, efficient local optimization alone is insufficient in the worst case.

1 INTRODUCTION

A foundational principle of computational complexity theory is that many important decisions are easier to verify than to make. Yet, even verification becomes difficult in practice once the relevant argument is too large for a human to inspect. The recursive debate protocol (Brown-Cohen et al. (2024); Irving et al. (2018)) addresses this problem.

At the start of a recursive debate, two agents disagree about the answer to a question that is too difficult for an external judge to resolve directly. The agent defending the current answer replaces it with several simpler subclaims and a local rule explaining how their answers determine the parent claim. The opponent chooses a subclaim on which to continue. Repeating this process keeps the judge's attention on a single path through a much larger argument until the debate reaches a claim that can be checked directly. This protocol is appealing because it appears to turn a large verification task into a sequence of small ones. If a decomposition faithfully represents its parent, then a false parent must leave some defect among its children. An honest opponent should be able to follow that defect downward, while an honest defender should be able to supply a consistent chain of increasingly simple claims.

The obfuscation problem. The missing step is finding the defective child. A dishonest defender can submit individually plausible subclaims whose inconsistency becomes visible only after a difficult global search. If the opponent follows the wrong branch, every claim eventually shown to the judge may look correct even though the original answer was false.

For example, suppose an agent claims that a long program execution is correct and decomposes this claim using intermediate program states. Checking one transition may be easy. Locating the incorrect intermediate state, however, may require reconstructing most of the execution. Recursing does not

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.

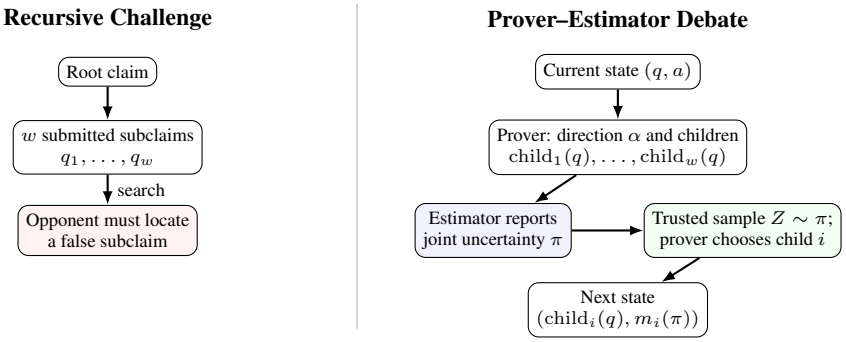

Figure 1: Recursive debate asks an opponent to find a false subclaim. In prover–estimator debate, the prover commits to a direction, the opponent reports uncertainty, and the prover selects a child after observing a trusted sample.

remove this burden; it gives the agent repeated opportunities to hide the failure. This is commonly referred to as the *obfuscated-arguments problem* Barnes et al. (2020).

Prover–estimator debate changes what the opponent must do (Brown-Cohen et al., 2026). The *prover* supplies the root answer and the recursive decompositions. The *estimator* is not asked to identify a false subclaim. Instead, it reports a joint probabilistic forecast for the submitted subclaims. The prover first declares whether it believes the current forecast is too high or too low. A trusted sample is drawn from the estimator's report, and the prover then chooses which subclaim to pursue. The chosen subclaim inherits the estimator's marginal forecast, and the process repeats.

The rewards are designed so that a prover can profit from a systematic inconsistency between the forecast for a parent claim and the forecasts for its children. The estimator can therefore defend itself without discovering the hidden error: it need only issue forecasts that remain coherent under whichever challenge the prover chooses. Trusted sampling ensures that the prover must expose an actual predictive inconsistency rather than merely assert that one exists.

Example 1.1 (Two-Rounds)**.** *Suppose the root claim is that all four unit tests in a software suite pass, while the actual outcomes are* $(1, 1, 1, 0)$*, where one denotes PASS and zero denotes FAIL. The fourth test fails, so the root claim is false. Further suppose that the estimator currently assigns the root claim probability of* $5/8$*. The prover declares this forecast too high and decomposes the root into two subclaims: "tests one and two both pass" and "tests three and four both pass." The estimator reports the joint forecast*

$$P_{root}(1,1) = 5/8, \quad P_{root}(1,0) = 2/8, \quad P_{root}(0,1) = 1/8$$

with all remaining probabilities equal to zero, where $(1, 0)$ *denotes that the first subclaim is true and the second is false. The marginals of this joint forecast are therefore* $7/8$ *and* $6/8$*.*

Now suppose that the trusted draw of outcomes is $(1, 0)$*, for which the sampled AND value is zero. In the absence of reward discounting, the prover's reward is the overestimation between the forecast and sampled value,* $5/8 - 0 = 5/8$*. The prover then continues with the second subclaim which, critically, inherits its marginal forecast of* $6/8$ *(not the sampled value of zero). The second subclaim asks whether tests three and four both pass. The prover again declares that the current forecast is too high and decomposes the claim into two individual tests. Suppose the estimator now reports:*

$$P_{\mathrm{pair}}(1,1) = 4/8, \quad P_{\mathrm{pair}}(1,0) = 2/8,$$
$$P_{\mathrm{pair}}(0,1) = 1/8, \quad P_{\mathrm{pair}}(0,0) = 1/8.$$

If the trusted draw is again $(1, 0)$*, the prover earns* $6/8 - 0 = 6/8$*. It continues with the fourth test, whose marginal forecast is* $5/8$*. We now have a trusted leaf check which returns zero, so the prover earns a final reward of* $5/8$*. The debate has thus followed one narrow path to the actual failure, although the estimator was never required to identify the failing test.*

As illustrated by the above example, an honest prover should be able to expose inaccurate forecasts by following a truthful recursive argument. Conversely, for any strategic prover, there should be an

efficient estimator whose forecasts cannot be systematically exploited. The second requirement must continue to hold for strategically chosen decompositions, as otherwise the hidden-search problem has merely moved into the estimator's forecasting task.

1.1 Our Contributions: the Constructive Question

The existence argument of (Brown-Cohen et al., 2026) updates the estimator using the true answers to the submitted queries and provides a circuit-size upper bound exponential in the width. Their work explicitly notes that this online-gradient construction is not executable, since each step requires access to the truth—its purpose is to establish existence. Our aim is complementary in that we give a uniform construction under additional, explicitly stated assumptions. In the present work, we ask:

> Can we efficiently construct, using only interaction with a fixed prover and tractable local optimization, an estimator whose forecasts the prover cannot systematically exploit?

Our work has three main contributions in service of this question.

1. **Efficient Estimator Construction.** Provided a submitted decomposition cannot profitably change the meaning of its parent claim, we reduce estimator construction to a one-sided local optimization problem that we call signed level-set optimization. The resulting compiler outputs a sound response to one fixed prover. It uses counterfactual queries to the fixed prover and simulated debate scores, not hidden true answers, and its construction and evaluation costs are polynomial in the system parameters.
2. **Weight-Threshold Instantiation.** We give an additive optimization oracle for binary-encoded weighted thresholds. Together with signed semantic admissibility, this yields an efficient sound estimator. Stable, efficiently implementable honest strategies additionally yield conditional equilibrium correctness.
3. **Total Search Hardness.** Without the semantic restriction, a uniform compiler for a promised class of instances would yield randomized polynomial-time algorithms for every polynomial-time verifiable total search relation, even when the local gates are projections or constants. Efficiently local optimization alone is therefore insufficient in the worst case.

2 Related work

Recursive and prover–estimator debate. Debate was proposed as a scalable-oversight mechanism by (Irving et al., 2018; Brown-Cohen et al., 2024) introduced doubly-efficient debate, in which honest strategies need only polynomially many steps even when the argument they defend is exponentially long. Our starting point is the prover–estimator protocol of Brown-Cohen et al. (2026), which replaces the task of identifying a false child with probabilistic prediction of the submitted children. That work proves existence through an online-gradient argument, which it explicitly describes as not executable, and motivates stability as the condition under which an honest prover can expose error with connections to fractional block sensitivity Kulkarni & Tal (2016). We retain its information order and recursive payoff, but study the critical algorithmic question: how can we implement the estimator without querying the true answers to the prover's children? We also isolate the semantic premise used when the source proof identifies a submitted decomposition with its parent. We give explicit semantic and local-optimization conditions that are sufficient for uniform construction with polynomial width dependence—conditions that carry their own accuracy costs.

Outcome indistinguishability and multicalibration. The estimator can be viewed as producing outcomes that should be indistinguishable from a hidden reference under tests induced by prover continuations. This connects the construction to outcome indistinguishability and its extensions beyond Bernoulli predictions (Dwork et al., 2021; 2022), as well as oracle-efficient multicalibration (Deng et al., 2023). The recursive setting adds three requirements not supplied directly by those frameworks: (1) tests may depend on the serialized report, (2) each report must have an explicit sparse realization, and (3) the exact marginal passed to the next round must come from the same joint distribution that is sampled by the verifier.

Optimization and total search. Our positive result uses projection-free sparse convex optimization (Frank & Wolfe, 1956; Jaggi, 2013) and a finite online-potential argument in the tradition of online convex optimization (Zinkevich, 2003; Hazan, 2016). The threshold oracle adapts standard knapsack

Table 1: Conceptual comparison. The gains in constructivity and width dependence rely on explicit semantic and computational assumptions.

	Brown-Cohen et al. (2026)	This work
Estimator Argument	Existence update using true submitted-child answers	Executable compiler using observed scores and signed level-set optimization
Width Dependence	Circuit size upper bound exponential in width	Polynomial, given efficient signed level-set optimization
Reports	General distribution on 2^w outcomes	At most $w + 2$ dyadic atoms
Semantics	Used inside the soundness telescope	Explicit sufficient semantic condition; hardness for a permissive class without this restriction

approximation ideas (Ibarra & Kim, 1975; Lawler, 1979). On the negative side, our compiler problem becomes a witness-extraction mechanism for arbitrary total relations, placing the obstruction in the TFNP framework of Megiddo & Papadimitriou (1991). This complements the position optimization result with a worst-case obstruction: the hard object can be the semantics of a refinement even when its Boolean gate is trivial.

3 Model & Problem Formulation

We now formalize the interaction described in the introduction. The model has two layers. A recursive proof system specifies how a claim is reduced to smaller claims. Prover–estimator debate places a game on top of that system: the prover proposes each reduction, while the estimator prices the uncertainty in the proposed children. Separating these layers is important because a syntactically legal reduction need not preserve the meaning of the parent.

3.1 Problem Instances & Recursive Decompositions

Definition 3.1 (Prover–Estimator Instance)**.** *A prover–estimator instance consists of a Boolean language L, an input distribution μ, a depth $d \geq 1$, a width $w \geq 1$, a discount parameter $\rho \geq 1$, a polynomial-time decomposition machine M_{dec}, and a Boolean terminal oracle $\mathcal{O}$. A level-k query q, for $k > 0$, may be refined into w level-$(k-1)$ queries. At a particular occurrence of q, the prover submits*

$$\mathrm{child}(q) := \big(\mathrm{child}_1(q), \ldots, \mathrm{child}_w(q)\big).$$

For proposed child answers $z \in \{0,1\}^w$, the resulting contextual gate is

$$M_q(z) := M_{\mathrm{dec}}\big(x, \mathrm{child}(q), z\big) \in \{0,1\}.$$

The root query is x. A level-zero query specifies a set S of bits in an implicit proof, and is evaluated as $\mathcal{O}(\omega_S)$.

The notation $\mathrm{child}_i(q)$ is transcript-local: it denotes the ith child submitted at this occurrence of q, not a canonical child of the encoded query. Query levels, types, and all public gate data are included in their encodings.

In the game, we consider two players as deterministic strategies A and B. Besides its interactive messages, a prover A exposes a context-free proof-bit interface $\omega^A(x)_j := A(x, j)$. These tagged calls receive no debate transcript, so they define one implicit proof string for each input. This provides consistency across calls, not correctness relative to the reference semantics introduced below. Estimator reports are canonical sparse dyadic distributions and hence the verifier can sample from them and compute their marginals exactly.

3.2 Game Interaction

Define $r := 1/\rho$. The recursive state is a pair (q, a), where q is the current query and $a \in [0,1]$ is the estimate inherited from the previous round. For a report $\pi \in \Delta(\{0,1\}^w)$, write $m_i(\pi) := \Pr_{Z\sim\pi}[Z_i = 1]$.

Definition 3.2 (Prover–Estimator Debate)**.** *On input x, the interaction is as follows.*

(i) Initialization. *The prover announces $A(x) \in \{0,1\}$ and the estimator reports $a = B(x)$. A trusted sample $Z^{\text{init}} \sim \text{Bern}(a)$ gives the prover reward $r_{\text{init}} = r^{d+1}\mathbf{1}\{Z^{\text{init}} = A(x)\}$. The recursive state begins at $(q, a) = (x, a)$.*

(ii) Internal round $k = d, \dots, 1$. *At state (q, a), the prover first announces a sign $\alpha \in \{-1, 0, 1\}$ and a refinement $\text{child}(q)$. The estimator reports π, the verifier samples $Z \sim \pi$, and the prover receives*

$$r_k = r^k \alpha\big(M_q(Z) - a\big).$$

After seeing Z, the prover chooses $i \in [w]$, and the next state is $(\text{child}_i(q), m_i(\pi))$.

(iii) Terminal round. *If the terminal state is (q, a) and q specifies S, the prover announces α_0 before the oracle call and receives $r_0 = \alpha_0\big(\mathcal{O}(\omega^A(x)_S) - a\big)$.*

The raw score and normalized payoff are defined to be

$$R := r_{\text{init}} + \sum_{k=0}^{d} r_k, \qquad H_{\rho,d} := r^{d+1} + \sum_{k=0}^{d} r^k, \qquad V := R/H_{\rho,d}.$$

We note the order of moves is critical to the game's structure: the prover commits to a direction before it sees the estimator's report, but chooses the next child after the trusted sample. Nevertheless, the next estimate is the report's unconditional marginal $m_i(\pi)$, not the sampled bit Z_i and not a probability conditioned on the prover's selection. This is precisely what allows errors on adjacent levels to cancel in expectation.

Throughout the work, we fix a legal prover A and seek a randomized compiler that uses the explicit description or resettable interfaces of A, samples from μ, and simulated debate scores, and returns a deterministic estimator B for which the expected score is small. The compiler may not query L or the true answers to the prover's submitted children. Thus, the goal is stronger than proving that a suitable estimator exists as the estimator must be constructible from information available in the interaction.

3.3 Semantic Assumptions

Local optimization alone cannot say whether an arbitrary submitted refinement means the same thing as its parent. We therefore state the semantic and computational assumptions separately

Assumption 3.3 (Total Consistent Reference Semantics)**.** *There is a function $g(x, q) \in \{0, 1\}$ on every legal query, with $g(x, x) = L(x)$. On every valid honest recursive proof, g agrees with the prescribed ground-truth answer. Its value is completely determined by x and the encoded query.*

At an internal level k, let $t^{(k)} := g(x, q^{(k)})$ and $z^{\star(k)} := \big(g(x, \text{child}_i(q^{(k)}))\big)_{i=1}^{w}$, and define the semantic residual as $\text{res}^{(k)} := M_{q^{(k)}}(z^{\star(k)}) - t^{(k)}$. At level zero, set $\text{res}^{(0)} := \mathcal{O}(\omega^A(x)_S) - g(x, q^{(0)})$.

Definition 3.4 (Signed Semantic Admissibility)**.** *A deterministic prover is signed semantically admissible if, on every legal transcript consistent with that prover and at every level, $\alpha^{(k)} \text{res}^{(k)} \le 0$.*

This condition isolates the minimal semantic requirement needed for local estimator guarantees to compose globally. A refinement need not preserve its parent exactly; it may have a nonzero semantic residual. What is ruled out is only a residual that helps the prover in its announced discrepancy direction. This one-sided requirement is exactly what the soundness telescope needs: semantic mismatches contribute nonpositively to the prover's payoff, without requiring the estimator to know the hidden reference answers.

For the computational assumption, let C be the complete legal context before an estimator move, including the sign α and contextual gate M_C. For $\alpha \neq 0$, define the estimator-favorable output value $\nu_\alpha = (1 - \alpha)/2$ and its level set

$$F_C^\alpha := \{z \in \{0,1\}^w : M_C(z) = \nu_\alpha\}.$$

Definition 3.5 (Efficient Signed Level-Set Optimization)**.** *A contextual gate family admits efficient signed level-set optimization if one deterministic algorithm, given C, rational costs c with $\|c\|_1 \le 1$, and $\varepsilon \in (0, 1]$, recognizes exactly whether F_C^α is empty and, when it is nonempty, returns $z \in F_C^\alpha$ satisfying $c \cdot z \le \min_{z' \in F_C^\alpha} c \cdot z' + \varepsilon$ in time polynomial in the encoded input lengths and $1/\varepsilon$.*

The preceding assumptions control the estimator-construction side of the game. To state completeness, we need one additional property of the gates along an honest recursive argument.

Definition 3.6 (Local Stability)**.** *Let $M : \{0,1\}^w \to \{0,1\}$ be a contextual gate and let $z^* \in \{0,1\}^w$ be its truthful child-answer vector. The gate is ρ-stable at z^* if, for every distribution π over proposed child-answer vectors,*

$$\Pr_{Z\sim\pi}\left[M(Z) \neq M(z^*)\right] \leq \rho \max_{i\in[w]} \Pr_{Z\sim\pi}\left[Z_i \neq z_i^*\right].$$

An honest recursive decomposition is ρ-stable if this inequality holds at the truthful child-answer vector of every gate that can be reached while following the honest decomposition.

The source proof implicitly needs semantic control at two points: it identifies the decomposition evaluated on the true submitted-child answers with the true parent answer, and it needs an analogous signed terminal inequality (Brown-Cohen et al., 2026). Assumption 3.3 and Definition 3.4 characterize this, but do not assert that the intended applications of the source fail. Rather, they identify the additional invariant those applications must supply. Section 4.3 shows that without such a restriction, uniform estimator construction is hard in the worst case for a permissive class of refinements.

4 Main Results

We give sufficient conditions for efficient estimator construction and show that efficient local optimization alone is insufficient. Under the assumptions of Section 3, signed local optimization is enough to compile a sound estimator. Threshold gates satisfy the optimization requirement at both signs and therefore give a weighted-threshold instantiation of the computational assumption. If the condition of Definition 3.4 is removed, even projection and constant gates can encode atrbitrary total search in the worst case.

4.1 Constructive Soundness Theorem

We first show that signed local optimization is sufficient to compile a sound estimator.

Definition 4.1 (Finite encoding and compiler access)**.** *All query and transcript encodings have polynomial length, and terminal sets are explicitly enumerable in polynomial time. Reports are canonical dyadic distributions with exact sampling and marginals; the prover sees the serialized report and the draw, not the sampler's random tape. The compiler receives independent samples from μ, complete-score simulation including terminal oracle calls, and resettable access to A: evaluation of its deterministic transcript-to-message functions on supplied legal, including counterfactual, histories. An explicit circuit suffices; a single nonresettable interaction or observed scores alone do not. It cannot query g or L. All costs, including output evaluation, are charged.*

Theorem 4.2 (Constructive Soundness)**.** *Under Assumption 3.3, fix a deterministic legal prover A satisfying Definition 3.4. Assume Definitions 3.5 and 4.1. For every rational $\eta \in (0,1]$ and $\beta \in (0,1)$, a randomized polynomial-time compiler outputs, with probability at least $1-\beta$, a deterministic estimator B satisfying*

$$\mathbb{E}_{x\sim\mu}\left[R^{A,B}(x)\right] \leq r^{d+1} \Pr_{x\sim\mu}\left[A(x) = L(x)\right] + \eta.$$

Every internal report is a dyadic distribution supported on at most $w+2$ atoms, with exact sampling and marginals.

The theorem constructs a sound response to one fixed prover, it does not minimize payoff to a specified approximation, construct one estimator effective against all provers, or compute an equilibrium. The benchmark $r^{d+1}\Pr[A(x) = L(x)]$ is the expected initialization reward under a truthful root forecast, although the compiler has no access to L or the submitted children's reference answers. In particular, an always-incorrect prover earns expected raw score at most η. The estimator need not recover those hidden answers: it only needs to prevent the fixed prover from exploiting its forecasts. No stability assumption is required for this construction; stability enters only through completeness below. The intermediate behavioral estimator, which uses private random indices, satisfies the expected-score bound pointwise for every input x, and the selected deterministic estimator satisfies it only in expectation over $x \sim \mu$, since score-based seed selection gives no simultaneous per-input guarantee. The construction resource bounds are polynomial in the public encoding lengths and relevant interface costs.

4.2 Threshold Gates and Equilibrium Correctness

A threshold gate has the form $M_{a,\tau}(z) = \mathbf{1}\left\{\sum_{i=1}^{w} a_i z_i \geq \tau\right\}$, where the nonnegative integer weights and threshold are binary encoded. Optimization over either output class reduces to a knapsack style covering problem.

Theorem 4.3 (Threshold-Gate Sound Estimator)**.** *Assume Assumption 3.3. Retain the fixed-prover, semantic admissibility, encoding, and access hypotheses of Theorem 4.2, and require the threshold parameters to be computable in polynomial time from every legal context. If every legal contextual gate is a polynomial-time computable threshold with binary-encoded nonnegative integer weights, then for every η and $\beta \in (0,1)$, the compiler outputs, with probability at least $1-\beta$, a deterministic estimator B with*

$$\mathbb{E}_{x\sim\mu}\left[R^{A,B}(x)\right] \leq r^{d+1} \Pr_{x\sim\mu}\left[A(x) = L(x)\right] + \eta.$$

Its reports have support at most $w+2$.

Observe that the threshold oracle does not yet establish semantic admissibility or supply an honest strategy. Stability only enters on the completeness side.

Definition 4.4 (Uniform honest strategy)**.** *Assume uniform polynomial-time routines supply the correct root answer, an implicit proof, and truthful children and their answers on every reachable honest continuation, for every $x \in \operatorname{supp}\mu$. All refinements and terminal evaluations agree with g, and every child permits an honest continuation. The strategy A_{hon} uses these routines, announces $\alpha = 2g(x,q) - 1$ at every level, and selects the smallest maximizer of $|m_i(\pi) - g(x, \mathrm{child}_i(q))|$, independently of the sample.*

Under Definition 4.4, the honest prover follows the child with largest marginal error and, with reward discount $r = 1/\rho$, its local rewards telescope to a global completeness guarantee. Specifically, for a ρ-stable honest decomposition, every legal estimator B and every $x \in \operatorname{supp}\mu$ satisfy $\mathbb{E}[R^{A_{\mathrm{hon}},B}(x)] \geq r^{d+1}$.

Definition 4.5 (Approximate A-leading Stackelberg equilibrium)**.** *For strategy classes $\mathcal{A}, \mathcal{B}$, write $\mathcal{V}(A,B) := \mathbb{E}_{x\sim\mu}[R^{A,B}(x)]/H_{\rho,d}$. An ε-approximate equilibrium consists of $A^* \in \mathcal{A}$ and a response map $B^* : \mathcal{A} \to \mathcal{B}$ satisfying $\mathcal{V}(A, B^*(A)) \leq \inf_{B\in\mathcal{B}} \mathcal{V}(A,B) + \varepsilon$ for every $A \in \mathcal{A}$ and $\mathcal{V}(A^*, B^*(A^*)) \geq \sup_{A\in\mathcal{A}} \mathcal{V}(A, B^*(A)) - \varepsilon$.*

Corollary 4.6 (Threshold debate at equilibrium)**.** *Fix $\eta \in (0,1]$ and strategy classes $\mathcal{A}, \mathcal{B}$. Suppose the hypotheses of Theorem 4.3 hold for every $A \in \mathcal{A}$, and $\mathcal{B}$ contains, for each such prover, a deterministic response satisfying the theorem's soundness bound at raw-score slack η. Assume also that $\mathcal{A}$ contains the polynomial-time honest strategy A_{hon} of Definition 4.4, whose honest decomposition is ρ-stable. Then, for every $\varepsilon \geq 0$, every ε-approximate A-leading Stackelberg equilibrium (A^*, B^*) (Definition 4.5) satisfies*

$$\Pr_{x\sim\mu}\left[A^*(x) \neq L(x)\right] \leq \frac{\eta + 2\varepsilon H_{\rho,d}}{r^{d+1}}.$$

add a bit about parametrization that gives ρ^{d+1}/ε?Thus, soundness supplies a response to each fixer prover, while completeness supplies one honest prover that succeeds against every legal estimator.

4.3 Hardness without Semantic Control

Let $\mathcal{Q}(x,u)$ be any polynomial-time verifiable total relation with witness length s. Totality means $\forall x\, \exists u \in \{0,1\}^{s(|x|)} : \mathcal{Q}(x,u) = 1$, where s is polynomially bounded and computable; padding ensures $s \geq 2$. We construct a prover that decodes a candidate witness W_B from the estimator's marginals. Coordinate projections propagate the decoded bits, and a final constant gate checks $\mathcal{Q}(x, W_B)$—Section 5.3 gives the construction. The compilation task requires polynomial-time construction and output evaluation, with success probability at least $2/3$.

Theorem 4.7 (Total-search hardness without semantic admissibility)**.** *Consider the promised task of compiling, for a fixed prover and point-mass input distribution, a deterministic estimator with raw expected score at most η, when some polynomial-size zero-score estimator is promised to exist. If this task has a randomized polynomial-time solution at $\eta = 1/3$, then $\mathsf{TFNP} \subseteq \mathsf{FBPP}$. This holds even for the constant-one language, constant-zero terminal oracle, and contextual gates that are only*

coordinate projections or constants, hence globally 1*-stable with exact signed level-set optimization. The reduction permits either* $(w, d) = (2, s)$ *or* $(w, d) = (s, 2)$.

The inclusion means that every polynomial-time verifiable total search relation has a randomized polynomial-time witness finder succeeding with probability at least $2/3$ on every input. This is a worst-case obstruction for a permissive class, not a characterization of all efficiently compilable instances or a proof that every violation of signed admissibility causes hardness.

5 Proof Architecture & Technical Overview

This section gives the proof architecture and isolates the role of each assumption in proving our main results. The complete arguments are deferred to the appendices due to space constraints.

5.1 Compiler and Signed Relaxation

The announced sign allows the estimator to control the prover's reward without matching the parent-gate moment exactly. At context C, introduce a surrogate gate bit ζ and the signed body

$$K_C^\alpha := \operatorname{conv}\big\{(z, \zeta) \in \{0,1\}^{w+1} : \alpha(M_C(z) - \zeta) \le 0\big\}.$$

For $\alpha = +1$, the defining inequality is $M_C(z) \le \zeta$ and for $\alpha = -1$, it is $M_C(z) \ge \zeta$. Thus, replacing the sampled gate output by the surrogate can only increase the prover's calculated reward, giving an upper bound suitable for soundness. Signed semantic admissibility places the truthful child vector paired with the parent reference answer inside this body (Lemma B.3). This pair is used only as a comparator in the analysis; the compiler never evaluates either reference answer.

Linear optimization over the signed body requires at most one call to the favorable-output level-set oracle: one branch is unconstrained, and the other fixes the surrogate to the estimator-favorable output (Lemma B.1). This can weaken the computational requirement relative to full-graph optimization. For example, CNF gates at sign $+1$ and DNF gates at sign -1 admit exact signed optimization, whereas full-graph optimization for CNFs at constant additive accuracy below $1/2$ decides satisfiability (Proposition B.2). This separation concerns the specified gate–sign combinations, not unrestricted CNF or DNF gates at both signs. For weighted thresholds, both output classes are tractable, so full-graph optimization is also tractable. As such, the signed relaxation additionally accommodates the weaker semantic condition.

At each context, the compiler runs a finite online optimization procedure over K_C^α. Frank & Wolfe (1956) updates produce explicit convex combinations of legal Boolean vertices; affine elimination and dyadic rounding preserve feasibility and leave at most $w + 2$ report atoms (Lemmas B.4 and B.5). The update direction is computable from the fixed prover's continuations. For each atom z of a candidate report π_t, the compiler evaluates the selected child $J_t(z)$ and the next announced sign $\alpha_t^-(z)$, passing that child its exact marginal forecast. These responses define the direction in Equation 4, which controls the current discrepancy together with the next round's signed marginal error. No reference-answer query or future estimator evaluation is needed.

The intermediate behavioral estimator publishes a uniformly selected iterate report. It does not publish the average report, which could induce different prover behavior. The online bound therefore applies even when the prover inspects the full serialized report (Lemma B.7). Exact marginal inheritance makes successive local error terms cancel, while signed semantic admissibility makes the remaining semantic residuals nonpositive (Proposition C.3). Finally, the compiler samples candidate seeds for the private iterate choices and compares their empirical complete-debate scores, selecting one deterministic estimator with high probability (Lemma C.4). Counterfactual prover queries drive the local updates while observed scores select the final seed.

The resulting construction is polynomial in inverse raw-score slack $1/\eta$. Translating this guarantee into target root error $\varepsilon_{\mathrm{root}}$ requires $\eta = \Theta(\varepsilon_{\mathrm{root}} r^{d+1})$, together with the equilibrium tolerance specified after Corollary 4.6. Its cost is therefore polynomial in $\rho^{d+1}/\varepsilon_{\mathrm{root}}$, which retains a dependence on depth through the reward discount.

5.2 Threshold Gate Machinery

Ordinary thresholds admit exact inference by sorting costs and scanning feasible prefix lengths (Lemma D.1). For weighted thresholds, constant cases are immediate; otherwise, output-one inference fixes negative-cost coordinates to one, leaving minimum-cost covering. Since $\|c\|_1 \le 1$,

rounding the remaining costs down in units of ε/w gives total rounded cost at most w/ε. A dynamic program stores the maximum attainable weight at each rounded cost, using $O(w^2/\varepsilon)$ updates on binary-encoded weights. Weights remain exact, preserving feasibility while losing at most ε in the objective. Complementing bits handles output zero using threshold $\sum_i a_i - \tau + 1$ and costs $-c$ (Lemma D.2).

Stability enters separately: an output flip requires a minimum weight of coordinate errors, so Markov's inequality gives a margin-dependent stability bound (Propositions D.4 and D.5). The honest prover announces the sign toward the true answer and selects the largest-error child. Under ρ-stability and $r = 1/\rho$,

$$\mathbb{E}[r_k] \geq r^k \mathbb{E}[e_k] - r^{k-1}\mathbb{E}[e_{k-1}],$$

where e_k denotes the inherited prediction error at level k. Terminal and initialization rewards complete the telescope to $\mathbb{E}R \geq r^{d+1}$ (Lemma E.1); comparing with soundness yields the equilibrium correctness bound (Corollary E.2). Expectations are over the trusted sampling randomness of the debate (and any private estimator randomness, if present).

5.3 Hardness via Marginal Decoding Channels

Fix a total relation $\mathcal{Q}$, an input x, and $\rho = 1$. The prover announces zero although $L \equiv 1$, uses sign -1 internally, always selects child one, and takes terminal sign zero with $\mathcal{O} \equiv 0$. It decodes W_B by thresholding estimator marginals at $1/2$: serially at $(w, d) = (2, s)$, or from one report at $(w, d) = (s, 2)$. Projection gates propagate estimates; the final submitted children encode W_B and induce the constant gate $\mathcal{Q}(x, W_B)$ (Lemma F.1). Exact marginal inheritance makes the sampling discrepancies conditionally mean zero. Including initialization, the rewards telescope in expectation (Lemma F.2):

$$\mathbb{E}_{x\sim\mu}\left[R^{A,B}(x)\right] = 1 - \Pr_{x\sim\mu}[\mathcal{Q}(x, W_B) = 1],$$

where the expectation and probability are over the trusted sampling randomness of a fresh debate (and any private randomness of a behavioral estimator), and compiler randomness is considered separately. This is not a pathwise identity: even deterministic estimators can adapt their reports to earlier trusted samples. Hardwiring any valid witness gives a polynomial-size zero-score estimator using point masses, establishing the promise (Lemma F.3). Conversely, expected score at most $1/3$ guarantees witness extraction with probability at least $2/3$ in a fresh, faithfully sampled debate. An efficient compiler, followed by repeated compilation and verified extraction, would therefore imply $\mathsf{TFNP} \subseteq \mathsf{FBPP}$ (Theorem F.4). A failed witness forces a positive signed semantic residual, locating the obstruction in semantic admissibility rather than local optimization (Proposition F.5).

6 Discussion & Implications

Our work was inspired by the UK AI Security Institute's Alignment Project which asks a number of questions of debate. The first of which asks when efficient estimator construction is possible and how its parameter dependence can be improved (UK AI Security Institute, n.d., Debate Subproblems 1 and 4). We give a conditional answer: signed semantic admissibility and efficient signed level-set optimization yield an executable compiler with polynomial width dependence. Weighted thresholds supply the optimization oracle, while sparse reports avoid enumerating all child outcomes. Our lower bound identifies a separate semantic obstruction: without admissibility, construction can encode arbitrary total search even with trivial gates and small zero-score estimators. This is a conditional hardness result for permissive refinements, not an unconditional exponential lower bound for sound proof systems. This directly connects to AISI's Debate Subproblem 5 on strategic and accidental mistakes (UK AI Security Institute, n.d.). Designing checkable refinement rules that enforce it is a concrete next target. Tractable inference does not itself supply stable honest arguments, as sought in AISI's stability question (Debate Subproblem 6). Threshold inference needs no margin, whereas suitable margins provide completeness. Moreover, target root error ε requires $\eta = \Theta(\varepsilon\rho^{-(d+1)})$ and corresponding equilibrium accuracy. Construction is therefore polynomial in ρ^{d+1}/ε, leaving unrestricted depth a challenge. Finally, responding to a fixed prover does not establish self-play convergence. Equilibrium correctness requires an implementable honest prover, and the analysis retains trusted sampling and terminal semantic control. AISI's exploration-hacking and systematic-human-error questions thus remain unresolved (Debate Subproblems 8 and 9), and empirical evaluations should therefore measure root correctness separately from resistance to exploitation by a fixed prover.

7 AI USE DISCLOSURE

We used ChatGPT 6 Pro and Gemini 3.1 Pro for language editing and technical proofreading as pre-review of mathematical statements and proofs, including identifying potential errors, gaps in arguments, and inconsistencies in notation or assumptions. The authors evaluated the models' suggestions and independently checked the final statements and proofs. The authors take responsibility for the correctness and originality of the paper.

REFERENCES

Beth Barnes, Paul Christiano, William Saunders, Joe Collman, Mark Xu, Chris Painter, Mihnea Maftei, and Ronny Fernandez. Debate update: Obfuscated arguments problem. In *AI Alignment Forum*, 2020.

Jonah Brown-Cohen, Geoffrey Irving, and Georgios Piliouras. Scalable AI safety via doubly-efficient debate. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, 2024. URL https://proceedings.mlr.press/v235/brown-cohen24a.html.

Jonah Brown-Cohen, Geoffrey Irving, Georgios Piliouras, Lijie Chen, Jiawei Li, and Zhiyang Xun. Avoiding obfuscation with prover–estimator debate. *arXiv preprint arXiv:2506.13609*, 2026. URL https://arxiv.org/abs/2506.13609. Version 2, revised July 2026.

Zhun Deng, Cynthia Dwork, and Linjun Zhang. HappyMap: A generalized multicalibration method. In *14th Innovations in Theoretical Computer Science Conference*, volume 251 of *LIPIcs*, pp. 41:1–41:23, 2023. doi: 10.4230/LIPIcs.ITCS.2023.41.

Cynthia Dwork, Michael P. Kim, Omer Reingold, Guy N. Rothblum, and Gal Yona. Outcome indistinguishability. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1095–1108, 2021. doi: 10.1145/3406325.3451064.

Cynthia Dwork, Michael P. Kim, Omer Reingold, Guy N. Rothblum, and Gal Yona. Beyond bernoulli: Generating random outcomes that cannot be distinguished from nature. In *Proceedings of the 33rd International Conference on Algorithmic Learning Theory*, volume 167 of *Proceedings of Machine Learning Research*, pp. 342–380, 2022. URL https://proceedings.mlr.press/v167/dwork22a.html.

Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1–2):95–110, 1956. doi: 10.1002/nav.3800030109.

Elad Hazan. *Introduction to Online Convex Optimization*, volume 2 of *Foundations and Trends in Optimization*. Now Publishers, 2016. doi: 10.1561/2400000013.

Oscar H. Ibarra and Chul E. Kim. Fast approximation algorithms for the knapsack and sum of subset problems. *Journal of the ACM*, 22(4):463–468, 1975. doi: 10.1145/321906.321909.

Geoffrey Irving, Paul Christiano, and Dario Amodei. AI safety via debate. *arXiv preprint arXiv:1805.00899*, 2018. URL https://arxiv.org/abs/1805.00899.

Martin Jaggi. Revisiting frank–wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 427–435, 2013. URL https://proceedings.mlr.press/v28/jaggi13.html.

Raghav Kulkarni and Avishay Tal. On fractional block sensitivity. *Chicago Journal of Theoretical Computer Science*, 2016(8):1–16, 2016. doi: 10.4086/cjtcs.2016.008.

Eugene L. Lawler. Fast approximation algorithms for knapsack problems. *Mathematics of Operations Research*, 4(4):339–356, 1979. doi: 10.1287/moor.4.4.339.

Nimrod Megiddo and Christos H. Papadimitriou. On total functions, existence theorems and computational complexity. *Theoretical Computer Science*, 81(2):317–324, 1991. doi: 10.1016/0304-3975(91)90200-L.

UK AI Security Institute. Computational complexity theory. The Alignment Project, n.d. URL https://alignmentproject.aisi.gov.uk/research-area/computational-complexity-theory. See "Complexity-Theoretic Approaches to Improving Debate Protocols". Accessed September 24, 2026.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*, pp. 928–936, 2003.

A Conventions and finite representations

We use w for the branching width and put

$$Q := w + 1, \qquad r := \rho^{-1}, \qquad S_{\rho,d} := \sum_{k=0}^{d} r^k, \qquad H_{\rho,d} := S_{\rho,d} + r^{d+1}.$$

We fix a public input length n and a distribution μ on $\{0,1\}^n$. A public polynomial upper bound on input length would suffice instead. All polynomial resource bounds are uniform in n, the public descriptions, w, d, the inverse target accuracy, and the logarithm of the inverse failure probability. Legal message and transcript length bounds are public and include the report precision. The report interface permits every canonical report of the chosen precision with at most $w + 2$ atoms, and every root estimate on that grid.

All numerical inputs are rational and binary encoded. Running-time bounds include bit operations, the costs of invoking the prover's interfaces, and the costs of writing oracle queries. In particular, polynomial time in d means polynomial in its numerical value. Throughout the soundness argument, the prover A is a fixed deterministic legal strategy. Its interactive interfaces may depend on the entire public transcript.

At level k, write q_k for the current query, a_k for its inherited estimate, α_k for the announced sign, and π_k for the estimator's report. Let

$$t_k := g(x, q_k), \qquad z_i^{\star,k} := g(x, \mathrm{child}_i(q_k)).$$

After the trusted sample $Z^k \sim \pi_k$, the prover chooses J_k, and

$$q_{k-1} = \mathrm{child}_{J_k}(q_k), \qquad a_{k-1} = m_{J_k}(\pi_k), \qquad t_{k-1} = z_{J_k}^{\star,k}.$$

The last equality uses the fact that reference answers depend on the encoded query, not its history. In particular, a_{k-1} is not the sampled bit and is not conditioned on the event that J_k was selected.

The raw score is

$$R = r^{d+1}\mathbf{1}\{Z^{\mathrm{init}} = A(x)\} + \sum_{k=1}^{d} r^k \alpha_k \big(M_{C_k}(Z^k) - a_k\big) + \alpha_0\big(\mathcal{O}(\omega^A(x)_S) - a_0\big),$$

where $Z^{\mathrm{init}} \sim \mathrm{Bernoulli}(a_d)$ and C_k is the complete context before the level-k report. Thus $|R| \le H_{\rho,d} \le d + 2$. The initialization sample is public before the first internal sign is announced. Terminal proof-bit calls are context-free: $\omega^A(x)_j = A(x, j)$.

A canonical report at precision b is a lexicographically ordered list of distinct atoms $z^1, \dots, z^s \in \{0,1\}^w$ with positive integer weights $v_1, \dots, v_s$ summing to 2^b. It represents $\pi = \sum_{\ell=1}^{s} v_\ell 2^{-b} \delta_{z^\ell}$. Sampling uses a uniform b-bit integer and cumulative integer weights; only the resulting atom is revealed. Its marginals are exactly

$$m_i(\pi) = 2^{-b} \sum_{\ell=1}^{s} v_\ell z_i^\ell.$$

Root estimates belong to the same dyadic grid. We fix deterministic tie-breaking in every algorithm below.

For soundness, the reference function is defined on every legal query, with $g(x, x) = L(x)$, and agrees with all valid honest recursive proofs. Signed semantic admissibility means that, on every legal transcript consistent with A, including transcripts using arbitrary legal estimator reports,

$$\alpha_k\big(M_{C_k}(z^{\star,k}) - t_k\big) \le 0 \quad (1 \le k \le d), \qquad \alpha_0\big(\mathcal{O}(\omega^A(x)_S) - t_0\big) \le 0.$$

These conditions do not require $A(x) = L(x)$.

The compiler receives an explicit circuit for A, or access to its tagged deterministic interfaces, independent samples from μ, and the ability to simulate complete debates and observe their scores, including terminal oracle calls. It has no access to g or L. Legal queries, reports, and transcripts have polynomial encoding length, and terminal subsets are explicitly enumerable in polynomial time. Output evaluation and all oracle-query costs are charged.

Interface access means evaluation of the fixed deterministic transcript-to-message functions on supplied syntactically legal histories, including counterfactual histories. Calls are stateless: querying one history neither changes the response to another nor adds messages to the actual debate. An explicit prover circuit supplies this access directly. For an interactive implementation it must be supplied by reset-and-replay access with all replay costs charged. Access only to one nonresettable live trajectory or to observed scores does not supply this interface. The prover's input consists of its public debate history; neither the compiler's seed nor the estimator's program is provided as an additional input.

For totality of the returned estimator, it may first check a supplied history against A's earlier messages and return the legal point mass δ_{0^w} on a mismatch. This check uses polynomially many calls to A on prefixes and requires no reference answers. Alternatively, with transcript-function access on all legal inputs, malformed counterfactual child choices or signs are replaced by child one or sign zero, respectively. Every report is then still obtained from legal lifted vertices. These defaults are never used in the soundness analysis, whose histories are consistent with the legal prover A. Out-of-format inputs may return δ_{0^w}, or root estimate zero.

B Signed geometry and finite online optimization

B.1 The signed body and its linear oracle

Fix a legal context C with announced sign α. Define

$$\mathcal{U}_C^\alpha := \big\{(z,\zeta) \in \{0,1\}^{w+1} : \alpha(M_C(z) - \zeta) \le 0\big\}, \qquad K_C^\alpha := \operatorname{conv}(\mathcal{U}_C^\alpha).$$

These sets are nonempty: $(z, M_C(z))$ is always a legal vertex. For $\alpha \neq 0$, put $\nu_\alpha = (1-\alpha)/2$ and $F_C^\alpha = M_C^{-1}(\nu_\alpha)$.

The assumed signed level-set oracle is one deterministic algorithm that, for every legal C with $\alpha \neq 0$, rational cost vector c with $\|c\|_1 \le 1$, and rational $\varepsilon \in (0,1]$, recognizes exactly whether F_C^α is empty and otherwise returns $z \in F_C^\alpha$ with $c \cdot z \le \min_{z' \in F_C^\alpha} c \cdot z' + \varepsilon$. Its running time is polynomial in the encoded inputs and $1/\varepsilon$. The guarantee is uniform over legal contexts, not only honest ones.

Lemma B.1 (Linear optimization over the signed body)**.** *Efficient signed level-set optimization gives the following oracle. Given rational $\ell \in \mathbb{Q}^{w+1}$ with $\|\ell\|_1 \le 1$ and $\varepsilon \in (0,1]$, it returns $v \in \mathcal{U}_C^\alpha$ satisfying*

$$\ell \cdot v \le \min_{u \in K_C^\alpha} \ell \cdot u + \varepsilon.$$

It uses at most one signed level-set optimization call and polynomial additional work. At sign zero it is exact and needs no such call.

Proof. Write $\ell = (c, h)$. For $\alpha \neq 0$ the legal vertices are exactly

$$\big(\{0,1\}^w \times \{1-\nu_\alpha\}\big) \cup \big(F_C^\alpha \times \{\nu_\alpha\}\big).$$

On the first branch, choose $z_i = \mathbf{1}\{c_i < 0\}$ and add the constant $h(1-\nu_\alpha)$ to its cost. On the second branch, use the signed oracle with costs c and accuracy ε, adding $h\nu_\alpha$. The second branch is discarded precisely when the oracle recognizes that F_C^α is empty. Return the cheaper candidate, with fixed tie-breaking. Its cost is at most the smaller of the two branch optima plus ε. A linear functional has the same minimum over a finite vertex set and its convex hull. For $\alpha = 0$, the body is the full $(w+1)$-dimensional cube and each coordinate is minimized independently. □

The announced sign can make the required oracle strictly easier than optimization over the full graph $\operatorname{conv}\{(z, M_C(z)) : z \in \{0,1\}^w\}$, which is contained in K_C^α.

Proposition B.2 (Positive CNF and negative DNF gates)**.** *Polynomial-size CNF gates at sign $+1$ and polynomial-size DNF gates at sign -1 admit exact polynomial-time signed level-set optimization, including recognition of an empty favorable class.*

Proof. For a positive-sign CNF, the favorable output is zero. A false assignment falsifies at least one clause. Enumerate the clauses; for each non-tautological clause, fix all its literals to false and optimize the remaining unary costs independently. Return the cheapest candidate. Every false assignment belongs to one of these subcubes, so this procedure is exact. If there is no falsifiable clause, the favorable class is empty. Empty clauses are handled as subcubes with no fixed coordinates. For a negative-sign DNF, the favorable output is one. Enumerate the consistent terms, fix their literals to true, and again optimize all remaining variables independently. An empty term imposes no constraints; if no consistent term exists, the favorable class is empty. Both procedures run in polynomial time in the formula and cost encodings. □

By contrast, an additive full-graph optimizer for CNFs at accuracy less than $1/2$, applied to $\min_z -M_C(z)$, returns a satisfying assignment whenever one exists and hence decides CNF satisfiability. The proposition concerns signed gate families: at sign -1 a CNF, or at sign $+1$ a DNF, would require optimization over satisfying assignments of a CNF (respectively, falsifying assignments of a DNF), so unrestricted CNF/DNF gates with both signs do not satisfy the same easy-oracle claim. Conversely, if favorable-class optimization is available at both signs of the same gate, as for weighted thresholds, the full graph can be optimized by optimizing each output class and comparing the two objective values.

Lemma B.3 (The reference comparator and reported moments)**.** *Suppose A is signed semantically admissible. At every legal context consistent with A, the vector*

$$v_C^\star := (z^\star, t), \qquad t = g(x, q), \quad z_i^\star = g(x, \mathrm{child}_i(q)),$$

belongs to $\mathcal{U}_C^\alpha$. Furthermore, if a convex combination of legal vertices has mean $u = (m, \bar{\zeta})$ and π is its projection onto the z coordinates, then $m = m(\pi)$ and

$$\alpha\big(\mathbb{E}_{Z\sim\pi} M_C(Z) - \bar{\zeta}\big) \le 0.$$

Proof. The first assertion is exactly $\alpha(M_C(z^\star) - t) \le 0$, which is signed admissibility at C. For the second assertion, take the weighted average of $\alpha(M_C(z) - \zeta) \le 0$ over the legal vertices in the representation. Projection preserves the first w coordinates of its mean. □

The comparator is used only in the analysis. Neither the oracle nor any update below evaluates g, t, or $z^\star$.

B.2 Sparse rational representations and dyadic rounding

Lemma B.4 (Constructive affine elimination)**.** *Given a rational convex combination of N Boolean vectors in $\{0,1\}^p$, one can compute an equal-mean convex combination supported on at most $p+1$ of those vectors in polynomial bit complexity.*

Proof. Merge duplicate vertices and delete zero coefficients. If more than $p+1$ positive coefficients remain, some at most $p+2$ augmented columns $(v, 1) \in \mathbb{Z}^{p+1}$ are linearly dependent. Exact Gaussian elimination supplies a nonzero integer dependence h satisfying

$$\sum_i h_i v^i = 0, \qquad \sum_i h_i = 0.$$

Here h is extended by zero to the other columns. Since $h \neq 0$ and its coordinates sum to zero, it has both positive and negative entries. For the current positive weights λ_i, set

$$\theta := \min_{i:h_i>0} \frac{\lambda_i}{h_i}, \qquad \lambda_i' := \lambda_i - \theta h_i.$$

The new weights are nonnegative, have total mass one, and have the same mean. At least one positive weight vanishes. Repeating terminates after at most N eliminations.

For the bit bound, a dependence on at most $p+2$ augmented Boolean columns can be chosen with integer entries of $O(p\log(p+1))$ bits, by determinant bounds. If the weights have a common denominator D, one elimination can be carried out with a common denominator dividing $D|h_j|$ for a minimizing index j. Therefore each elimination adds only $O(p\log(p+1))$ denominator bits. Numerators have polynomial length as well, since the weights remain in $[0,1]$. The number of arithmetic operations and the lengths of their operands are polynomial. □

Lemma B.5 (Dyadic realization)**.** *Let $y=\sum_{i=1}^{s}\lambda_i v^i$ be a rational convex combination of Boolean vertices, with $s\le p+1$. In polynomial time one can round the weights to multiples of 2^{-b}, preserving their sum and using only these vertices, to obtain y' satisfying*

$$\|y'-y\|_\infty \le s2^{-b}.$$

Zero weights can be deleted. If $p\le Q$ and $e:=s2^{-b}\le 1$, then, for every $v\in[0,1]^p$,

$$\|y'-v\|_2^2 \le \|y-v\|_2^2 + 3Qe.$$

Proof. Initially assign integer weight $\lfloor 2^b\lambda_i\rfloor$ to vertex i. Distribute the remaining units to the largest fractional remainders, breaking ties by vertex order. Each individual probability changes by at most 2^{-b}, so the sum of absolute changes is at most $s2^{-b}$. Since every coordinate of every vertex is in $\{0,1\}$, the stated coordinate error follows.

Write $\Delta=y'-y$. Then $\|\Delta\|_2\le\sqrt{Q}e$ and $\|y-v\|_2\le\sqrt{Q}$. Consequently

$$\|y'-v\|_2^2-\|y-v\|_2^2 = 2(y-v)\cdot\Delta+\|\Delta\|_2^2 \le 2Qe+Qe^2\le 3Qe.$$

□

Applied to lifted vertices (z,ζ), these operations preserve membership in the signed body. After rounding, merge vertices with the same z coordinate when publishing the report. There are at most $Q+1=w+2$ published atoms. The lifted mean remains an internal quantity; only the canonical distribution over z is communicated.

B.3 A FINITE APPROXIMATE PROJECTION ROUTINE

We next give the precise finite optimization primitive used in the online updates. Its small starting objective is important for the stated running time.

Lemma B.6 (Sparse approximate projection)**.** *Let $K=\mathrm{conv}(\mathcal{U})\subseteq[0,1]^p$, where $p\le Q$ and $\mathcal{U}\subseteq\{0,1\}^p$ is nonempty. Suppose an additive linear-optimization oracle returns vertices of $\mathcal{U}$ for normalized rational costs. Fix $0<\delta\le 1$ and set*

$$\lambda:=\delta/8,\qquad \tau:=\lambda\delta/8.$$

Suppose $u\in K$ is given as a dyadic combination of at most $Q+1$ vertices and $D\in\mathbb{Q}^p$ satisfies $\|D\|_1\le 2$. If

$$(Q+1)2^{-b}\le\frac{\lambda\delta}{12Q}, \tag{1}$$

one can compute a dyadic combination $u^+\in K$ of at most $p+1$ vertices such that, simultaneously for every $v\in K$,

$$\|u^+-v\|_2^2\le\|u-\lambda D-v\|_2^2+\frac{\lambda\delta}{2}. \tag{2}$$

The routine makes $O(Q/\delta^2)$ oracle calls, each at normalized accuracy $\Theta(\delta^2/Q)$, and has polynomial bit complexity.

Proof. Put $a=u-\lambda D$ and minimize $f(y)=\frac{1}{2}\|y-a\|_2^2$ over K, starting from $y=u$. At the current y, the gradient is $g=y-a$. Since $y,u\in[0,1]^p$,

$$\|g\|_1\le Q+2\lambda\le Q+2.$$

Call the linear oracle with costs $g/(Q+2)$ and accuracy $\tau/[2(Q+2)]$. Its returned vertex s satisfies

$$g\cdot s\le\min_{v\in K} g\cdot v+\tau/2.$$

Define the computable gap $\widehat{G} = g \cdot (y - s)$. If $\widehat{G} \le \tau/2$, stop. Otherwise update

$$y \leftarrow (1-\gamma)y + \gamma s, \qquad \gamma := \frac{\tau}{2Q}.$$

The step size is fixed and rational. It belongs to $(0, 1]$.

At a nonterminating iteration, quadratic expansion gives

$$f((1-\gamma)y + \gamma s) = f(y) - \gamma\widehat{G} + \frac{\gamma^2}{2}\|s - y\|_2^2 < f(y) - \frac{\tau^2}{8Q},$$

because $\|s - y\|_2^2 \le Q$ and $\widehat{G} > \tau/2$. Initially $f(u) = \lambda^2\|D\|_2^2/2 \le 2\lambda^2$, whereas $f \ge 0$. Thus there are at most $16Q\lambda^2/\tau^2$ nonterminating updates, which is $O(Q/\delta^2)$.

At termination, the true variational gap obeys

$$\max_{v \in K} g \cdot (y - v) \le \widehat{G} + \tau/2 \le \tau.$$

For every $v \in K$ this yields

$$\|y - v\|_2^2 = \|a - v\|_2^2 - \|y - a\|_2^2 + 2g \cdot (y - v) \le \|a - v\|_2^2 + 2\tau.$$

The iterates are explicit convex combinations of the starting support and the returned vertices. Apply Lemma B.4, then Lemma B.5. By equation 1, rounding adds at most $\lambda\delta/4$ to squared distance from any $v \in K$. Also $2\tau = \lambda\delta/4$, proving equation 2.

For completeness, exact rational arithmetic does not introduce an uncontrolled precision cost. Before compression, each update multiplies weights by the same rational constants γ and $1 - \gamma$. After N updates their denominator lengths grow by at most $O(N \operatorname{bits}(\gamma))$ beyond the initial representation. Lemma B.4 gives a polynomial bit bound for compression, and rounding resets the output to the fixed precision b. All oracle inputs consequently have polynomial encoding length. □

B.4 Online regret without reference-answer access

Lemma B.7 (Finite online bound)**.** *Under the preceding oracle assumptions, initialize at any Boolean vertex $u_1 \in K$. At time t, after u_t and its chosen canonical representation are fixed, let an arbitrary rule produce $D_t \in \mathbb{Q}^p$ with $\|D_t\|_1 \le 2$. Update by Lemma B.6. If $T \ge 8Q/\delta^2$, then every $v \in K$ satisfies*

$$\frac{1}{T}\sum_{t=1}^{T} D_t \cdot (u_t - v) \le \delta. \tag{3}$$

The rule producing D_t may inspect the full serialized report; no continuity, differentiability, or dependence only on moments is required.

Proof. Equation equation 2 gives

$$\begin{aligned}\|u_{t+1} - v\|_2^2 &\le \|u_t - v\|_2^2 - 2\lambda D_t \cdot (u_t - v) + \lambda^2\|D_t\|_2^2 + \lambda\delta/2 \\ &\le \|u_t - v\|_2^2 - 2\lambda D_t \cdot (u_t - v) + 4\lambda^2 + \lambda\delta/2.\end{aligned}$$

Summing and using $\|u_1 - v\|_2^2 \le Q$ gives

$$\frac{1}{T}\sum_{t=1}^{T} D_t \cdot (u_t - v) \le \frac{Q}{2\lambda T} + 2\lambda + \delta/4 \le \delta/2 + \delta/4 + \delta/4 = \delta.$$

The argument uses only the realized vectors D_t and their norm bounds, so arbitrary dependence on the current representation is allowed. The unknown comparator is never needed to compute an update. □

C Constructive soundness: the compiler and its proof

C.1 Computable local feedback

We first define a behavioral estimator, meaning an intermediate strategy with private random choices. Those choices will subsequently be fixed by a polynomial-time compiler. Fix a local error parameter δ, a power of two $T \ge 8Q/\delta^2$, and a precision satisfying equation 1.

Internal contexts. At a fixed context C of level $k \geq 1$, run the online procedure on K_C^α, initialized at $(0^w, M_C(0^w))$. Write its current lifted mean as

$$u_t = (m_t, \bar{\zeta}_t),$$

and let π_t be the canonical dyadic report obtained by projecting its current representation onto the child coordinates.

For every atom z of π_t, simulate the prover's legal continuation from the fixed history C after publishing π_t and revealing z. Let $J_t(z)$ be its selected child. Pass the exact estimate $m_{t,J_t(z)}$ to that child, and obtain the next announced sign $\alpha_t^-(z)$. When $k = 1$, this is the terminal sign; no terminal oracle call is needed to determine it. Define

$$\beta_{t,i} := \sum_{z \in \operatorname{supp} \pi_t} \pi_t(z)\alpha_t^-(z)\mathbf{1}\{J_t(z) = i\}, \qquad D_t := (-\beta_t, r\alpha). \tag{4}$$

The sum is computed exactly from at most $w + 2$ atoms. In particular,

$$\|D_t\|_1 \leq \sum_z \pi_t(z)|\alpha_t^-(z)| + r|\alpha| \leq 2.$$

Use this feedback to compute u_{t+1}. All these calls hold the history before the report fixed; they are counterfactual evaluations of the deterministic prover, not extra messages in the actual debate. Every queried successor is legal and has positive probability under the corresponding report. The procedure never queries the child reference answers.

The behavioral estimator chooses a uniform index $I \in [T]$, privately, and publishes π_I. It does not publish an average of the reports. The full local sequence is deterministic given C and can equivalently be generated only as far as the selected index.

Lemma C.1 (One internal cancellation inequality)**.** *At every legal context C consistent with a signed-admissible prover, the preceding behavioral report satisfies*

$$\mathbb{E}\left[r_k - r^k\alpha_k(t_k - a_k) + r^{k-1}\alpha_{k-1}(t_{k-1} - a_{k-1}) \mid C\right] \leq r^{k-1}\delta. \tag{5}$$

The expectation is over the private uniform index and the ensuing trusted sample. The next sign is evaluated before the next estimator move, or before the terminal oracle call.

Proof. At this fixed context, $\vartheta = t_k$, $z^\star$, $a = a_k$, and $\alpha = \alpha_k$ are fixed. Lemma B.3 makes $v^\star = (z^\star, \vartheta)$ a feasible comparator. For a particular iterate,

$$\begin{aligned} D_t \cdot (u_t - v^\star) &= r\alpha(\bar{\zeta}_t - \vartheta) + \sum_i \beta_{t,i}(z_i^\star - m_{t,i}) \\ &= r\alpha(\bar{\zeta}_t - \vartheta) + \mathbb{E}_{Z\sim\pi_t}[\alpha_t^-(Z)(z^\star_{J_t(Z)} - m_{t,J_t(Z)})]. \end{aligned}$$

By Lemma B.3,

$$\mathbb{E}_{Z\sim\pi_t}[r_k] \leq r^k\alpha(\bar{\zeta}_t - a) = r^k\alpha(\vartheta - a) + r^k\alpha(\bar{\zeta}_t - \vartheta).$$

Moreover, exact inheritance and fixed reference semantics identify $z^\star_{J_t(Z)} - m_{t,J_t(Z)}$ with $t_{k-1} - a_{k-1}$. Multiplying the previous inner-product identity by r^{k-1} therefore bounds the conditional left side of equation 5 for that iterate. Average over the uniform index and apply Lemma B.7 with comparator $v^\star$. □

The root. On input x, write $b_A = A(x)$. Run the same online procedure on $[0, 1]$, initialized at zero, with scalar iterate p_t. For each $z \in \{0, 1\}$ of positive probability under $\mathrm{Bernoulli}(p_t)$, simulate the first internal sign after the public root report p_t and initialization sample z. Denote it by $\alpha_{d,t}(z)$. Set

$$D_t^{\mathrm{root}} := r(2b_A - 1) - \mathbb{E}_{Z\sim\mathrm{Bernoulli}(p_t)}\alpha_{d,t}(Z). \tag{6}$$

This rational scalar has absolute value at most two. Choose a uniform private index and report the corresponding p_t. The scalar projection routine can use Q as an upper bound on its dimension and diameter squared, so the common T and b remain sufficient.

Lemma C.2 (Initialization cancellation)**.** *For every input x, the behavioral root strategy satisfies*

$$\mathbb{E}\left[r_{\text{init}} + r^d \alpha_d(t_d - a_d) \mid x\right] \le r^{d+1}\mathbf{1}\{A(x) = L(x)\} + r^d\delta. \tag{7}$$

Proof. Put $t = L(x)$. For a fixed root iterate p,

$$\begin{aligned}\mathbb{E}[r_{\text{init}} \mid p, x] - r^{d+1}\mathbf{1}\{b_A = t\} \\ = r^{d+1}(2b_A - 1)(p - t).\end{aligned}$$

Adding $r^d\mathbb{E}[\alpha_d(t - p) \mid p, x]$ gives exactly $r^d D^{\text{root}}(p - t)$. The bit t is a feasible comparator in $[0, 1]$. Averaging the online bound over the uniform root index proves equation 7. □

C.2 Telescoping the raw score

Proposition C.3 (Behavioral soundness)**.** *Use independent private uniform indices at initialization and at each internal level. Under total reference semantics and signed semantic admissibility, the resulting behavioral estimator B_{beh} satisfies, for every input x,*

$$\mathbb{E}[R^{A,B_{\text{beh}}}(x)] \le r^{d+1}\mathbf{1}\{A(x) = L(x)\} + S_{\rho,d}\delta. \tag{8}$$

Proof. Set $E_k := r^k\alpha_k(t_k - a_k)$. Taking expectations in Lemma C.1 gives, for each $k \ge 1$,

$$\mathbb{E}[r_k - E_k + E_{k-1}] \le r^{k-1}\delta.$$

At the terminal level, signed admissibility gives

$$r_0 = E_0 + \alpha_0\big(\mathcal{O}(\omega^A(x)_S) - t_0\big) \le E_0.$$

The exact identity

$$R = (r_{\text{init}} + E_d) + \sum_{k=1}^{d}(r_k - E_k + E_{k-1}) + (r_0 - E_0)$$

now telescopes. Lemma C.2 bounds its first term. The total error is

$$r^d\delta + \sum_{k=1}^{d} r^{k-1}\delta = S_{\rho,d}\delta.$$

All local inequalities hold conditional on their actual pre-report contexts, so the tower property applies even when earlier reports and trusted samples determine those contexts adaptively. □

Notice that this proof never replaces a selected marginal by a conditional probability. Also, averaging strategies is essential: an averaged report could induce different child selections and signs, so the online bound does not justify substituting that report.

C.3 Fixing randomness using observable scores

For a seed

$$\sigma = (I_{\text{init}}, I_d, \ldots, I_1) \in [T]^{d+1},$$

let B_σ be the deterministic estimator that uses its designated index at each level. Each level is visited once in a debate. Thus a uniform seed, chosen independently of x and all trusted samples, has exactly the same single-debate distribution as the behavioral strategy. Indeed, the local sequence at a context uses only that context and calls to A; it never reads an index assigned to a later level. Inductively, the history before a level depends only on earlier indices and trusted samples. Conditional on that history, the unused index remains uniform and independent, exactly as in the behavioral construction. This argument is about a randomly chosen seed before score-based selection, not about the distribution of the selected seed. The seed length is $(d + 1)\log_2 T$ bits. In particular,

$$\mathbb{E}_\sigma\mathbb{E}_{x\sim\mu,\,\text{trusted coins}}R^{A,B_\sigma}(x) \le b_0 + S_{\rho,d}\delta, \qquad b_0 := r^{d+1}\Pr_{x\sim\mu}[A(x) = L(x)]. \tag{9}$$

The compiler does not know b_0.

Lemma C.4 (Score-based deterministic selection)**.** *Suppose a samplable family of deterministic estimators satisfies $|R| \le H$, where $H \ge 1$, and*

$$\mathbb{E}_\sigma F(\sigma) \le b_0 + \eta/4, \qquad F(\sigma) := \mathbb{E}R^{A,B_\sigma},$$

where $0 \le b_0 \le 1$ and $0 < \eta \le 1$. Given complete-score simulation, one can select a seed with $F(\sigma) \le b_0 + \eta$, with probability at least $1 - \beta$, without knowing b_0. It suffices to sample

$$M = \left\lceil \frac{8H}{\eta} \log \frac{2}{\beta} \right\rceil$$

independent seeds and evaluate each on

$$N = \left\lceil \frac{32H^2}{\eta^2} \log \frac{4M}{\beta} \right\rceil$$

independent complete debates, returning a seed with smallest empirical mean score.

Proof. Let $p = \Pr_\sigma[F(\sigma) \le b_0 + \eta/2]$. If $b_0 + \eta/2 > H$, then $p = 1$. Otherwise, since $F(\sigma) \ge -H$,

$$b_0 + \eta/4 \ge \mathbb{E}F(\sigma) \ge -Hp + (b_0 + \eta/2)(1 - p).$$

Hence $p \ge \eta/(8H)$. In both cases, the probability that none of the M candidates satisfies $F \le b_0 + \eta/2$ is at most $\exp(-M\eta/(8H)) \le \beta/2$.

For a fixed candidate, independent scores in $[-H, H]$ have empirical mean $\widehat{F}$ satisfying

$$\Pr[|\widehat{F} - F| > a] \le 2 \exp\left(-\frac{Na^2}{2H^2}\right).$$

One elementary justification is the bounded-variable exponential moment bound. For $X \in [-H, H]$, put $\phi(t) = \log \mathbb{E}e^{t(X - \mathbb{E}X)}$. Under the exponentially tilted law, $\phi''(t)$ is the variance of a variable supported on an interval of length $2H$, and is at most H^2. Since $\phi(0) = \phi'(0) = 0$, integration gives $\phi(t) \le t^2H^2/2$. Applying Markov's inequality to the exponential of the sum and optimizing t proves the upper tail; replacing X by $-X$ proves the lower tail. The variance bound follows, for example, from $\mathbb{E}(X - a)(b - X) \ge 0$ on an interval $[a, b]$.

Apply the concentration inequality at $a = \eta/4$ and take a union bound over the candidates. The chosen N makes the probability of any inaccurate empirical mean at most $\beta/2$, conditional on the sampled seeds and hence also unconditionally. On the two successful events, there is a candidate σ_0 with $F(\sigma_0) \le b_0 + \eta/2$, and the selected seed $\widehat{\sigma}$ satisfies

$$F(\widehat{\sigma}) \le \widehat{F}(\widehat{\sigma}) + \eta/4 \le \widehat{F}(\sigma_0) + \eta/4 \le F(\sigma_0) + \eta/2 \le b_0 + \eta.$$

Selection therefore uses only scores, not the unknown benchmark. For a purely rational implementation, choose any integer upper bounds on the displayed sample sizes: replacing each natural logarithm by the ceiling of its base-two logarithm is sufficient and can be done by integer comparisons. This changes neither the guarantees nor the polynomial resource bounds. □

C.4 The complete compiler theorem

Theorem C.5 (Constructive soundness, with resource bounds)**.** *Assume total reference semantics, signed semantic admissibility of a fixed deterministic legal prover A, polynomially bounded legal encodings, the compiler access specified in the main text, and efficient signed level-set optimization. For every rational $\eta \in (0, 1]$ and $\beta \in (0, 1)$, a randomized polynomial-time compiler outputs, with probability at least $1 - \beta$, a deterministic estimator B satisfying*

$$\mathbb{E}_{x \sim \mu,\, \text{trusted coins}} R^{A,B}(x) \le r^{d+1} \Pr_{x \sim \mu}[A(x) = L(x)] + \eta.$$

Every internal report has at most $w + 2$ dyadic atoms, with exact sampling and exact marginal inheritance. One can take

$$T = O\left(\frac{(w+1)S_{\rho,d}^2}{\eta^2}\right), \qquad b = O\left(\log \frac{(w+2)S_{\rho,d}}{\eta}\right).$$

With prover-interface access only, the output is a deterministic A-oracle circuit; with an explicit circuit for A, its calls can be inlined. No oracle for g or L is used.

Proof. Set

$$\delta = \frac{\eta}{4S_{\rho,d}}, \qquad \lambda = \frac{\delta}{8},$$

choose T to be the least power of two at least $8Q/\delta^2$, and choose

$$b = \left\lceil \log_2 \frac{96Q(Q+1)}{\delta^2} \right\rceil.$$

This precision satisfies equation 1. Use the signed-body oracle of Lemma B.1, the finite projection routine, and the local feedback algorithms above. Equation equation 9 bounds the seed-average score by $b_0 + \eta/4$. Apply Lemma C.4 with $H = H_{\rho,d}$, using independent samples from μ and complete debate simulations. Its conclusion is the required inequality.

We verify the computational claim explicitly. A local online sequence has $T = O(Q/\delta^2)$ iterations. Each projection makes $O(Q/\delta^2)$ signed-oracle calls at accuracy $\Theta(\delta^2/Q)$. Consequently, one internal report requires at most

$$O(Q^2/\delta^4) = O(Q^2 S_{\rho,d}^4/\eta^4)$$

signed-oracle calls. Feedback at each outer iterate enumerates at most $w + 2$ report atoms and uses polynomially many arithmetic operations and prover-interface calls. It invokes no later estimator, so this calculation has no hidden recursion in its evaluation cost.

All intermediate arithmetic has polynomial bit complexity by Lemmas B.4–B.6. The quantities r^k, $S_{\rho,d}$, and $H_{\rho,d}$ have polynomial encoding lengths. Every actual or counterfactual report is canonical and has polynomial length; the assumed encoding bounds therefore apply to the histories passed to A. Evaluation of the returned estimator, not merely its construction, is polynomial in the stated parameters.

The numbers of candidate seeds and test debates are polynomial in $H_{\rho,d}$, $1/\eta$, and $\log(1/\beta)$. Complete-score simulation uses only the allowed terminal oracle and context-free proof-bit interface. Thus every required operation is available to the compiler. The finite support and exact-marginal claims follow from affine elimination, dyadic rounding of lifted vertices, and merging equal child atoms. Finally, the displayed asymptotic bounds on T and b follow from the explicit choices above. □

The deterministic conclusion is distributional: a single selected seed is guaranteed to work in expectation over μ. The behavioral inequality was pointwise in x, but score-based seed selection does not assert a simultaneous per-input guarantee. The theorem also does not claim exact payoff minimization against A.

D THRESHOLD GATES: OPTIMIZATION AND STABILITY

D.1 EXACT INFERENCE FOR ORDINARY THRESHOLDS

For an integer τ, let

$$M_\tau(z) := \mathbf{1}\left\{ \sum_{i=1}^{w} z_i \geq \tau \right\}.$$

The gate is constant when $\tau \leq 0$ or $\tau > w$.

Lemma D.1 (Exact ordinary-threshold inference)**.** *For rational costs $c \in \mathbb{Q}^w$ of either sign and a desired output $\nu \in \{0,1\}$, one can recognize whether $M_\tau^{-1}(\nu)$ is empty and otherwise find an exact minimum-cost vector in it using $O(w \log(w+1))$ rational comparisons and $O(w)$ further arithmetic operations. Its bit complexity is polynomial in the input length.*

Proof. Sort the coordinates so that $c_{(1)} \leq \cdots \leq c_{(w)}$, using original indices to resolve ties. Set $C_0 = 0$ and $C_j = \sum_{i=1}^{j} c_{(i)}$. Among vectors with exactly j ones, selecting the j cheapest coordinates minimizes the objective and gives value C_j. Therefore

$$\min_{M_\tau(z)=\nu} c \cdot z = \min_{\substack{0 \leq j \leq w \\ \mathbf{1}\{j \geq \tau\} = \nu}} C_j,$$

with the minimum over an empty set interpreted as $+\infty$. A scan finds a minimizing feasible cardinality or certifies emptiness. Returning its cheapest prefix proves the claim. Signed costs are allowed: a minimum need not have cardinality exactly τ. Sorting, rational addition, and the comparisons all have polynomial bit complexity. □

D.2 Additive inference for binary-encoded weighted thresholds

Let

$$M_{a,\tau}(z) := \mathbf{1}\left\{\sum_{i=1}^{w} a_i z_i \geq \tau\right\}, \qquad a_i \in \mathbb{Z}_{\geq 0}, \quad \tau \in \mathbb{Z}, \qquad W := \sum_i a_i.$$

The weights and threshold are binary encoded; their numerical magnitudes need not be polynomially bounded.

Lemma D.2 (Weighted-threshold inference)**.** *Given rational c with $\|c\|_1 \leq 1$, a desired output $\nu \in \{0,1\}$, and rational $\varepsilon \in (0,1]$, a deterministic algorithm recognizes exactly whether $M_{a,\tau}^{-1}(\nu)$ is empty. If it is nonempty, it returns an exactly feasible vector z such that*

$$c \cdot z \leq \min_{M_{a,\tau}(z')=\nu} c \cdot z' + \varepsilon.$$

The algorithm uses $O(w^2/\varepsilon)$ dynamic-programming updates on integers of polynomial bit length, plus polynomial preprocessing. Its running time is polynomial in the encoded inputs and $1/\varepsilon$, not in the numerical value of W.

Proof. If $\tau \leq 0$, the gate is constantly one; if $\tau > W$, it is constantly zero. In either case the desired class is either empty or the entire cube, and independent coordinate minimization suffices. Hence assume $1 \leq \tau \leq W$, so both output classes are nonempty.

First require output one. Let $z_i^0 = \mathbf{1}\{c_i < 0\}$ be an unconstrained minimizer. Return it if it is feasible. Otherwise put

$$D := \tau - \sum_i a_i z_i^0 > 0, \qquad I := \{i : c_i \geq 0\}.$$

An optimal feasible vector can be chosen with all negative-cost coordinates equal to one, since turning any such coordinate on both preserves the covering constraint and decreases cost. Thus the remaining task is to select $S \subseteq I$ with $\sum_{i\in S} a_i \geq D$ while minimizing $\sum_{i\in S} c_i$. It is feasible because $\tau \leq W$.

Set $h = \varepsilon/w$ and round only the costs:

$$\ell_i := \lfloor c_i/h \rfloor \quad (i \in I), \qquad K := \sum_{i\in I} \ell_i \leq w/\varepsilon.$$

Enumerate $I = \{i_1, \ldots, i_m\}$. For $0 \leq j \leq m$ and $0 \leq v \leq K$, let $\mathsf{D}_j(v)$ be the maximum weight attainable by a subset of the first j items whose rounded cost is exactly v. Unattainable entries are $-\infty$. Initialize $\mathsf{D}_0(0) = 0$ and $\mathsf{D}_0(v) = -\infty$ for $v > 0$, and use

$$\mathsf{D}_j(v) = \max\left\{\mathsf{D}_{j-1}(v), \quad a_{i_j} + \mathsf{D}_{j-1}(v - \ell_{i_j})\right\},$$

omitting out-of-range entries. Both terms use the previous row, even when $\ell_{i_j} = 0$. Store predecessors with deterministic tie-breaking. Return a reconstructible subset S for the smallest v with $\mathsf{D}_m(v) \geq D$.

Let $S^\star$ minimize the original covering cost. The chosen rounded cost is at most that of $S^\star$, so

$$\begin{aligned}\sum_{i\in S} c_i &\leq h \sum_{i\in S} \ell_i + wh \\ &\leq h \sum_{i\in S^\star} \ell_i + \varepsilon \leq \sum_{i\in S^\star} c_i + \varepsilon.\end{aligned}$$

Adding the fixed cost of the negative coordinates proves the desired approximation. Feasibility is exact because neither the weights nor the covering threshold was rounded.

For output zero, integrality gives

$$\sum_i a_i z_i \leq \tau - 1 \quad \Longleftrightarrow \quad \sum_i a_i (1 - z_i) \geq W - \tau + 1.$$

Use variables $x_i = 1 - z_i$ and costs $-c_i$, whose ℓ_1 norm is still at most one. The original objective is $c \cdot z = \sum_i c_i + (-c) \cdot x$, so applying the output-one algorithm and complementing its answer preserves the additive error.

There are $O(w^2/\varepsilon)$ table entries. Each stored finite weight is at most W and therefore needs only $O(\log(W+1))$ bits. Scaling, comparisons, and reconstruction have polynomial bit complexity. □

Exact weighted inference would be stronger than what is needed here. For example, when $1 \le \tau \le W$, the optimum of

$$\min_{M_{a,\tau}(z)=1} \sum_i (a_i/W) z_i$$

is τ/W precisely when some subset has weight exactly τ. The additive algorithm need not decide this equality at inverse-polynomial accuracy, since successive achievable objective values can be only $1/W$ apart.

Theorem D.3 (Threshold-gate sound estimator)**.** *Retain the reference semantics, signed admissibility, finite encoding, and compiler access assumed in Theorem C.5. Suppose every legal context supplies, in polynomial time, binary-encoded nonnegative integer weights and an integer threshold of polynomial encoding length. Then the conclusions and the support bound of that theorem hold. No margin or stability assumption is required for this conclusion.*

Proof. Lemma D.2 supplies signed level-set optimization at both signs, with exact feasibility and correct empty-set recognition. Its guarantee is uniform over all legal contexts, including ones a dishonest prover may submit. Apply Theorem C.5. For ordinary thresholds, Lemma D.1 supplies an exact oracle instead.

More explicitly, the generic compiler requests local inference accuracy $\Theta(\eta^2/(QS_{\rho,d}^2))$ and makes $O(Q^2 S_{\rho,d}^4/\eta^4)$ such calls per internal report. The weighted implementation therefore uses

$$O\left(\frac{w^2 Q^3 S_{\rho,d}^6}{\eta^6}\right)$$

dynamic-programming table updates per report, besides the other polynomial costs already bounded above. This is a bound for the elementary implementation, not an optimal accuracy dependence. □

D.3 Local stability of thresholds

For any distribution π and truthful Boolean vector $z^\star$, put

$$e := \max_i \Pr_{Z\sim\pi}[Z_i \ne z_i^\star] = \max_i |m_i(\pi) - z_i^\star|.$$

The equality uses that each reference coordinate is a bit. Likewise, for a Boolean gate,

$$\Pr[M(Z) \ne M(z^\star)] = |\mathbb{E}M(Z) - M(z^\star)|.$$

These identities make the probability and marginal formulations of local stability equivalent.

Proposition D.4 (Exact ordinary-threshold stability)**.** *Let $1 \le \tau \le w$ and $h = \sum_i z_i^\star$. The optimal local stability constant is*

$$\rho_\tau(z^\star) = \begin{cases} \dfrac{h}{h-\tau+1}, & h \ge \tau, \\ \dfrac{w-h}{\tau-h}, & h < \tau. \end{cases}$$

The Hamming distance to the opposite output class is

$$D_\tau(z^\star) = \begin{cases} h-\tau+1, & h \ge \tau, \\ \tau-h, & h < \tau. \end{cases}$$

Thus $D_\tau(z^\star) \ge \gamma w$ implies local stability at most $1/\gamma$. Constant gates have stability zero, allowing zero as an optimal constant in this statement.

Proof. Suppose $h \geq \tau$. Let X count original one-coordinates that turn to zero in Z. An output flip requires $X \geq h - \tau + 1$, even if some original zeros also turn to one. Since $\mathbb{E}X \leq he$, Markov's inequality gives

$$\Pr[M_\tau(Z) = 0] \leq \frac{he}{h - \tau + 1}.$$

For tightness, choose a uniformly random $(h - \tau + 1)$-subset of the h original one-coordinates and turn exactly these coordinates off. The output flips with probability one, while each original one has error probability $(h - \tau + 1)/h$ and every other coordinate has error zero. No smaller stability constant is possible.

If $h < \tau$, let X count original zero-coordinates turned on. An output flip requires $X \geq \tau - h$ and $\mathbb{E}X \leq (w - h)e$. This proves the second upper bound. Flipping a uniformly random $(\tau - h)$-subset of the original zeros proves its tightness. The same minimum-flip arguments give the two Hamming distances. Each stability numerator is at most w, proving the margin implication. □

Proposition D.5 (Weighted-margin stability)**.** *Assume $1 \leq \tau \leq W$ and put $h = \sum_i a_i z_i^\star$. A valid local stability constant is*

$$\overline{\rho}_{a,\tau}(z^\star) = \begin{cases} \dfrac{h}{h - \tau + 1}, & h \geq \tau, \\ \dfrac{W - h}{\tau - h}, & h < \tau. \end{cases}$$

Consequently, the relevant margin condition

$$h - \tau + 1 \geq \gamma W \quad (h \geq \tau), \qquad \tau - h \geq \gamma W \quad (h < \tau)$$

implies local stability at most $1/\gamma$. The displayed upper bound need not be tight for unequal weights.

Proof. For $h \geq \tau$, let $X = \sum_{i:z_i^\star = 1} a_i \mathbf{1}\{Z_i = 0\}$ be the weight of original ones that turn off. An output flip requires $X \geq h - \tau + 1$, since all weights are integers. Also $\mathbb{E}X \leq he$. Markov's inequality gives the first bound. For $h < \tau$, use $X = \sum_{i:z_i^\star = 0} a_i \mathbf{1}\{Z_i = 1\}$, with $\mathbb{E}X \leq (W - h)e$ and a flip requiring $X \geq \tau - h$. Each numerator is at most W, which proves the margin implication. No independence between coordinate errors was used. □

E Completeness and equilibrium correctness

E.1 The honest strategy and the completeness telescope

Lemma E.1 (Completeness from local stability)**.** *Suppose uniform honest routines, on every input in* $\operatorname{supp}\mu$*, return the correct root answer, supply a valid implicit proof, and at every reachable honest query produce legal children and all their truthful answers. Assume that every selected child admits an honest continuation, every honest refinement satisfies $M_C(z^\star) = g(x, q)$, and terminal proof-bit evaluations equal the reference answer. If all these honest gates are ρ-stable at their truthful vectors, there is one honest strategy A_{hon} such that, for every legal estimator B and every such input x,*

$$\mathbb{E}[R^{A_{\mathrm{hon}},B}(x)] \geq r^{d+1}.$$

Its overhead beyond the honest routines is polynomial in the report length and w. In particular, polynomial-time honest routines yield a polynomial-time honest strategy; existence of such routines is an additional hypothesis, not a consequence of the soundness compiler.

Proof. At each level let $e_k = |a_k - t_k|$. The honest strategy announces $\alpha_k = 2t_k - 1$, so $\alpha_k(t_k - a_k) = e_k$. It supplies the honest refinement, reads the report's exact marginals, and chooses

$$J_k \in \arg\max_{i \in [w]} |m_i(\pi_k) - z_i^{\star,k}|,$$

using the smallest maximizing index. Although this move occurs after the sample, its rule is independent of the sampled vector. Exact marginal inheritance gives

$$e_{k-1} = \max_i |m_i(\pi_k) - z_i^{\star,k}|.$$

Let $\mathcal{H}_k$ be the sigma-field generated by the complete history through the report, before its trusted draw. Both e_k and e_{k-1}, as well as C_k and t_k, are $\mathcal{H}_k$-measurable. Truthful refinement and stability imply

$$|\mathbb{E}[M_{C_k}(Z^k) \mid \mathcal{H}_k] - t_k| \le \rho e_{k-1}.$$

Consequently,

$$\mathbb{E}[r_k \mid \mathcal{H}_k] \ge r^k e_k - r^k \rho e_{k-1} = r^k e_k - r^{k-1} e_{k-1}.$$

The genuine terminal check gives $r_0 = e_0$. The correct root announcement gives expected initialization reward $r^{d+1}(1 - \mathbb{E}e_d)$. Summing the inequalities therefore yields

$$\begin{aligned}\mathbb{E}R &\ge r^{d+1}(1 - \mathbb{E}e_d) + \sum_{k=1}^{d}(r^k \mathbb{E}e_k - r^{k-1}\mathbb{E}e_{k-1}) + \mathbb{E}e_0 \\ &= r^{d+1} + r^d(1-r)\mathbb{E}e_d \ge r^{d+1}.\end{aligned}$$

The strategy computes w marginal errors and invokes the assumed honest routines; the additional work is polynomial. Uniform access to these routines is a hypothesis: the mere existence of unrelated witnesses for individual inputs would not supply one implementable strategy over the distribution. □

E.2 Approximate Stackelberg equilibria

Write

$$\mathcal{V}(A, B) := \frac{\mathbb{E}_{x\sim\mu,\,\text{trusted coins}} R^{A,B}(x)}{H_{\rho,d}}.$$

For strategy classes $\mathcal{A}, \mathcal{B}$, an ε-approximate A-leading Stackelberg equilibrium consists of $A^* \in \mathcal{A}$ and a response map $B^* : \mathcal{A} \to \mathcal{B}$ such that

$$\begin{aligned}\mathcal{V}(A, B^*(A)) &\le \mathcal{V}(A, B) + \varepsilon \qquad \text{for every } A \in \mathcal{A},\ B \in \mathcal{B}, \\ \mathcal{V}(A^*, B^*(A^*)) &\ge \mathcal{V}(A, B^*(A)) - \varepsilon \quad \text{for every } A \in \mathcal{A}.\end{aligned}$$

The follower minimizes the prover's payoff.

Corollary E.2 (Equilibrium correctness)**.** *Suppose the soundness hypotheses hold for every $A \in \mathcal{A}$, $\mathcal{B}$ contains the corresponding compiled responses at raw slack η, and $\mathcal{A}$ contains the honest strategy of Lemma E.1. Then every such equilibrium satisfies*

$$\Pr_{x\sim\mu}[A^*(x) \ne L(x)] \le \frac{\eta + 2\varepsilon H_{\rho,d}}{r^{d+1}}.$$

Proof. Completeness and the leader inequality imply

$$\mathcal{V}(A^*, B^*(A^*)) \ge \mathcal{V}(A_{\text{hon}}, B^*(A_{\text{hon}})) - \varepsilon \ge \frac{r^{d+1}}{H_{\rho,d}} - \varepsilon.$$

For the single prover A^* choose a sound compiled response $B_{A^*} \in \mathcal{B}$. Such a response exists because the compiler succeeds with positive probability. Soundness and the follower inequality give

$$\mathcal{V}(A^*, B^*(A^*)) \le \mathcal{V}(A^*, B_{A^*}) + \varepsilon \le \frac{r^{d+1}\Pr[A^*(x) = L(x)] + \eta}{H_{\rho,d}} + \varepsilon.$$

Combining and rearranging proves the result. No simultaneous successful compilation over the entire prover class is needed. □

E.3 Accuracy, margins, and computational dependence

For target root error $\varepsilon \in (0, 1]$, it suffices to set

$$\eta = \frac{\varepsilon r^{d+1}}{2}, \qquad \varepsilon \le \frac{\varepsilon r^{d+1}}{4H_{\rho,d}}.$$

The compiler is therefore polynomial in ρ^{d+1}/ε, besides its other parameters. This dependence is distinct from the polynomial dependence on branching width.

Under either threshold margin condition with a common rational $\gamma \in (0, 1]$, one can choose $\rho = 1/\gamma$. If a size parameter $N \geq 2$ bounds the public encodings, w, d, and primitive costs, polynomial-cost construction at fixed correctness accuracy follows whenever $\gamma^{-(d+1)}$ is polynomial in N. Constant $\gamma > 0$ and $d = O(\log N)$ are a sufficient regime. The margin is imposed only on the honest argument, not on every threshold a strategic prover may submit. These conclusions neither construct an equilibrium nor eliminate the honest-routine and semantic hypotheses.

F TOTAL-SEARCH HARDNESS WITHOUT SEMANTIC ADMISSIBILITY

F.1 THE PROMISED CONSTRUCTION TASK

Let $\mathcal{Q}(x, u)$ be a polynomial-time predicate with polynomially bounded, computable witness length $\ell(|x|)$ and

$$\forall x \quad \exists u \in \{0,1\}^{\ell(|x|)} : \mathcal{Q}(x, u) = 1.$$

Such predicates specify the total search relations considered here. The notation $\mathsf{TFNP} \subseteq \mathsf{FBPP}$ means that every such relation has a randomized polynomial-time algorithm returning a valid witness with probability at least $2/3$ on every input.

The compiler receives a finite instance description, an explicit polynomial-time deterministic prover, a point-mass input distribution, and a raw slack η. It has the sample-and-simulate access of the positive theorem, but signed semantic admissibility is not promised. Instead, it is promised that a polynomial-size deterministic estimator B_0 with $\mathbb{E}R^{A,B_0} = 0$ exists. The task is to output, with probability at least $2/3$, a deterministic estimator B satisfying $\mathbb{E}R^{A,B} \leq \eta$. Construction, output size, and output evaluation must all be polynomial in the public parameters and $1/\eta$. The last requirement excludes programs that defer unbounded search until evaluation. These are worst-case resource bounds, including unsuccessful compilation outcomes. Evaluation of a returned program is clocked at the promised polynomial bound; timeout, malformed code, or an invalid report makes that compile-and-extract attempt fail. Equivalently, one may require the compiler always to return a legal, bounded-time estimator and allow only its score guarantee to fail.

We retain exact finite reports and the original move order. The refinement syntax below checks only the specified structural forms; it does not enforce preservation of the current query's reference answer. This is a permissive-refinement lower bound, not a lower bound for the signed-admissible class. The promise requires an existential honest tree with the correct root value, not a polynomial-time routine for producing one; the honest-routine hypothesis of completeness is not imposed in this lower-bound task.

F.2 THE COPY–CHECK INSTANCES

Fix a relation and an input x. Put

$$s := \max\{2, \ell(|x|)\}, \qquad \overline{\mathcal{Q}}(x, W) := \mathcal{Q}(x, W_{1:\ell(|x|)}), \quad W \in \{0,1\}^s.$$

Set $L \equiv 1$, $\rho = 1$, and $\mathcal{O} \equiv 0$. The implicit proof has length $m = s$, and all prover proof-bit answers will be zero. Let $\mathsf{H}(W)$ denote the terminal subset

$$S(W) := \{i \in [s] : W_i = 1\},$$

encoded by its s-bit indicator. Thus W is encoded in an actual terminal query, rather than an unbounded auxiliary message. Query encodings also specify their types and levels.

Serial instance. Take $(w, d) = (2, s)$. Interpret the root as carrying the empty prefix. At level $k \geq 2$, a parent carrying a prefix a of length $s - k$ may submit, for either $b \in \{0, 1\}$,

$$(\mathsf{C}(ab), \mathsf{C}(ab)),$$

where the copy query $\mathsf{C}(ab)$ is at level $k - 1$. The contextual gate on this list is $z \mapsto z_1$. At level one, a parent carrying a prefix a of length $s - 1$ may submit

$$(\mathsf{H}(ab), \mathsf{H}(ab)), \qquad b \in \{0, 1\},$$

whose contextual gate is the constant $z \mapsto \overline{\mathcal{Q}}(x, ab)$. The syntax verifies levels, duplication, and prefix extension, but not whether a prefix has a valid witness extension.

Parallel instance. Take $(w, d) = (s, 2)$. The root must submit the list of distinct level-one copy queries $(\mathsf{C}_1, \ldots, \mathsf{C}_s)$, whose gate is $z \mapsto z_1$. From any C_i, a level-one refinement may submit s copies of $\mathsf{H}(W)$ for any $W \in \{0,1\}^s$. Its gate is the constant $z \mapsto \overline{\mathcal{Q}}(x, W)$. Again, syntax checks this form without requiring the predicate to be true.

A single polynomial-time decomposition machine for either instance recognizes its copy lists and applies z_1, recognizes duplicated terminal subsets and applies $\overline{\mathcal{Q}}$ to their indicator, and returns zero on other lists. The syntax checker excludes the other lists during a legal execution.

Lemma F.1 (Elementary gates and an honest tree)**.** *Every contextual gate in these instances is a projection or a constant. It has global stability at most one and exact polynomial-time signed level-set optimization. Each input admits a recursively evaluated honest tree with root value one and implicit-proof length $m = s \le w^d$. All query encodings and primitive algorithms have polynomial size.*

Proof. For a projection, the probability of a gate error is exactly the error probability in its selected coordinate, hence is at most the maximum coordinate error. A constant gate never changes output. Optimizing unary costs on a projection's output class fixes its selected coordinate and optimizes all others independently. For a constant, the desired class is either empty or the whole cube. These gates are also weighted thresholds: use a single weight one and threshold one for the projection, or all weights zero with threshold zero or one for a constant.

By totality choose $W^\star$ with $\overline{\mathcal{Q}}(x, W^\star) = 1$. In the serial tree, label all nodes at distance $t \in \{1, \ldots, s-1\}$ from the root by $\mathsf{C}(W^\star_{1:t})$, and every terminal node by $\mathsf{H}(W^\star)$. Terminal answers are zero, the last internal gates return one independently of them, and all projection gates above also return one. The parallel tree has the prescribed copy queries, each expanded to terminal queries $\mathsf{H}(W^\star)$; the same bottom-up evaluation gives root value one.

The proof-length bounds are $s \le 2^s$ and $s \le s^2$, respectively. Only the succinct syntax and gate routines, not an expanded serial tree, are supplied to the compiler. Prefixes and terminal indicators have length at most s, and their structural checks and predicate evaluations are polynomial-time. □

F.3 The fixed prover and witness decoding

In both instances, let $A_{\mathcal{Q}}$ announce zero initially, use $\alpha_k = -1$ at every internal round, always select child $J_k = 1$, and use terminal sign $\alpha_0 = 0$. Its context-free proof bits are all zero. Since $L \equiv 1$, its initial announcement is always wrong. Write $P_k = a_k$.

In the serial instance, at the start of round k the prover computes

$$b_k := \mathbf{1}\{P_k \ge 1/2\}$$

and appends this bit to the current prefix. At copy rounds it submits the corresponding duplicated copy queries. At level one it submits the duplicated terminal subset for

$$W_B := b_d b_{d-1} \cdots b_1.$$

In the parallel instance, it first submits the required copy list. From the ensuing root report π_2, it decodes

$$(W_B)_i := \mathbf{1}\{m_i(\pi_2) \ge 1/2\}, \qquad i \in [s],$$

and at level one submits the duplicated terminal subset for W_B. The entire earlier report is part of the transcript, so the prover can recover these bits even though only $m_1(\pi_2)$ is inherited as the current estimate. It still selects child one after the sample. All these operations are polynomial-time and require no witness search.

Even for deterministic B, the variable W_B can be random: later reports may depend on previous trusted samples. The identities below retain that randomness.

Lemma F.2 (Exact expected-payoff identity)**.** *For either instance and every legal estimator B,*

$$\mathbb{E}R^{A_{\mathcal{Q}},B}(x) = 1 - \Pr[\overline{\mathcal{Q}}(x, W_B) = 1]. \tag{10}$$

The statement also holds for behavioral estimators. Since $\rho = 1$,

$$\mathbb{E}V^{A_{\mathcal{Q}},B}(x) = \frac{1 - \Pr[\overline{\mathcal{Q}}(x, W_B) = 1]}{d+2}.$$

Proof. The initial claim is zero, so $r_{\text{init}} = 1 - Z^{\text{init}}$. At every copy round $k \geq 2$,

$$r_k = P_k - Z_1^k, \qquad P_{k-1} = m_1(\pi_k).$$

At level one, the constant gate gives $r_1 = P_1 - \overline{\mathcal{Q}}(x, W_B)$, and $r_0 = 0$. Therefore the following identity holds on every execution:

$$R = 1 - \overline{\mathcal{Q}}(x, W_B) + (P_d - Z^{\text{init}}) + \sum_{k=2}^{d} (P_{k-1} - Z_1^k). \tag{11}$$

Conditional on the history through π_k, the selected child is already known to be one, and $P_{k-1} = m_1(\pi_k)$ is fixed. Hence

$$\mathbb{E}[P_{k-1} - Z_1^k \mid \text{history through } \pi_k] = 0.$$

The initialization discrepancy also has conditional mean zero. Taking expectations proves equation 10. No independence between W_B and the discrepancies is needed. Dividing by $H_{1,d} = d + 2$ gives the normalized identity. The failure indicator equals the raw score in expectation, not necessarily pathwise. □

Lemma F.3 (A polynomial-size zero-score estimator exists)**.** *Every constructed input admits a deterministic polynomial-size estimator B_0 whose raw score against $A_{\mathcal{Q}}$ is zero on every execution. Its root estimate is a bit and every internal report is a point mass.*

Proof. Hardwire any valid padded witness $W^\star$ for x. In the serial instance, set $P_d = W_1^\star$ and, at each copy round $k \geq 2$, report

$$\pi_k = \delta_{(W^\star_{s-k+2},\, W^\star_{s-k+2})}.$$

The successive inherited estimates are exactly the successive bits of $W^\star$, so the final submitted subset is $\mathsf{H}(W^\star)$. In the parallel instance, set $P_2 = 0$ and $\pi_2 = \delta_{W^\star}$. The marginal decoder again returns $W^\star$. In both instances, use δ_{0^w} for the final internal report.

Every sample on these executions is deterministic. Thus all zero-mean discrepancies in equation 11 vanish pathwise, as does $1 - \overline{\mathcal{Q}}(x, W^\star)$. Hence $R = 0$. Arbitrary legal point masses define the strategy on other histories. Hardwiring x, $W^\star$, and these rules requires polynomial circuit size. Point masses are legal at every positive precision, so the promise does not depend on fine probability encodings. □

F.4 The extraction reduction

Theorem F.4 (Total-search hardness)**.** *If the promised construction task has a randomized polynomial-time solution at raw slack $\eta = 1/3$, then $\mathsf{TFNP} \subseteq \mathsf{FBPP}$. This holds for the constant-one language, constant-zero terminal oracle, and only projection and constant gates. For padded witness length s, either $(w, d) = (2, s)$ or $(w, d) = (s, 2)$ suffices.*

Proof. Fix an arbitrary total relation and search input x. Construct either instance and the explicit prover $A_{\mathcal{Q}}$, use $\mu = \delta_x$, and invoke the assumed compiler at $\eta = 1/3$. Lemma F.3 verifies the existence promise; its hardwired witness is not given to the compiler.

Whenever compilation succeeds, the returned B satisfies, by Lemma F.2,

$$\Pr[\overline{\mathcal{Q}}(x, W_B) = 1 \mid B] \geq 2/3.$$

Simulate a fresh complete debate against B, sampling exactly from each report and passing the resulting actual history to every later estimator call. Apply the preceding resource clocks and format checks; on an invalid output, declare this attempt unsuccessful. Read W_B from the final submitted terminal subset, verify $\overline{\mathcal{Q}}(x, W_B)$, and return its first $\ell(|x|)$ bits only when the verification accepts. Determinism of B does not permit replacing its sampled history by an arbitrary fixed history.

A compile-and-extract attempt succeeds with probability at least $(2/3)(2/3) = 4/9$. Repeat the entire attempt independently three times, verifying each candidate. The failure probability is at most $(5/9)^3 < 1/3$.

Every operation in the reduction is polynomial-time. The input sampler always returns x, the terminal oracle always returns zero, and the public gates use only the efficiently computable predicate and

trivial projections. Thus every interface and score-simulation call available to the compiler can be implemented inside the search algorithm. The required output-evaluation bound makes fresh debate simulation polynomial-time as well. We have obtained a randomized polynomial-time witness finder for an arbitrary total relation. □

The normalized slack corresponding to raw slack $1/3$ is $1/[3(d+2)]$, which is inverse-polynomial in the serial construction and is $1/12$ in the parallel construction. The latter also has $w^d = s^2$ polynomially bounded. The two regimes do not assert a simultaneous constant-width, constant-depth lower bound.

F.5 Identifying the semantic violation

Proposition F.5 (Witness failure is a positive semantic residual)**.** *The constructed instances admit total reference semantics agreeing with their witness-consistent honest trees such that, on every execution of $A_{\mathcal{Q}}$,*

$$\sum_{k=0}^{d} \alpha_k \operatorname{res}^{(k)} = 1 - \overline{\mathcal{Q}}(x, W_B).$$

In particular, every failed decoded witness entails a violation of signed semantic admissibility.

Proof. For a serial prefix a, define

$$T_x(a) := \mathbf{1}\{\exists v \in \{0,1\}^{s-|a|} : \overline{\mathcal{Q}}(x, av) = 1\}.$$

Assign value $T_x(a)$ to $\mathsf{C}(a)$, value one to the root, and value zero to each terminal query. Totality gives $T_x(\varnothing) = 1$. For a full string W, $T_x(W) = \overline{\mathcal{Q}}(x, W)$, although its terminal subset $\mathsf{H}(W)$ still has oracle value zero. These assignments are functions of encoded queries and agree with the honest trees built from valid witnesses.

Let $a_0 = \varnothing, a_1, \ldots, a_s = W_B$ be the chosen prefixes. At a copy round the residual is $T_x(a_{j+1}) - T_x(a_j)$ because both child references equal the longer prefix's value. At the final check it is the same difference because the constant gate returns $T_x(a_s)$. Every internal sign is -1, so

$$\sum_{k=1}^{d} \alpha_k \operatorname{res}^{(k)} = \sum_{j=0}^{s-1} (T_x(a_j) - T_x(a_{j+1})) = 1 - \overline{\mathcal{Q}}(x, W_B).$$

The terminal residual is zero. A prefix that destroys all valid completions creates a positive signed residual.

In the parallel instance, assign reference value one to the root and all copy queries and zero to terminal queries. The root projection has zero residual; the final check has residual $\overline{\mathcal{Q}}(x, W_B) - 1$ and sign -1. This gives the same identity. No algorithm for evaluating the prefix-existence predicate is supplied to the compiler; these values only identify the semantic obstruction. □

The theorem separates uniform construction from small-estimator existence. It does not assert that checking an estimator's expected score is a polynomial-time predicate, that the compiler task itself belongs to TFNP, or that the positive theorem applies to these permissive provers.

G Further conditional and oracle consequences

Corollary G.1 (One-way permutations)**.** *Suppose there is a uniformly polynomial-time computable family of bijections $f_n : \{0,1\}^n \to \{0,1\}^n$ such that every randomized polynomial-time algorithm inverts f_n on a uniformly random image with only negligible probability. Then the compiler described in Theorem F.4 cannot exist.*

Proof. Use $\mathcal{Q}_f(x, u) = \mathbf{1}\{f_{|x|}(u) = x\}$ with $|u| = |x|$. Bijectivity makes this relation total, and polynomial-time evaluation of f_n makes it efficiently verifiable. The theorem would produce an inverter succeeding with probability at least $2/3$ on every image, contradicting the assumption. □

This is a conditional standard-model exclusion, not an unconditional exponential lower bound. An exponential statement is available in a classical forward-oracle model, counting compilation and subsequent evaluation together.

Lemma G.2 (Forward-query inversion bound)**.** *Let Π be a uniformly random permutation of $\{0,1\}^n$, let U be uniform and independent of Π, and reveal $x = \Pi(U)$. A classical randomized algorithm making at most $q_0 < 2^n$ distinct forward queries to Π, with no advice depending on Π, satisfies*

$$\Pr[\Pi(\widehat{U}) = x] \le \frac{q_0 + 1}{2^n}.$$

Proof. Put $N_0 = 2^n$ and first condition on the algorithm's private random tape. Given a query transcript that has not returned x, the preimage of x is uniform over the unqueried domain points, by symmetry of the random permutation. After j distinct unsuccessful queries, the next fresh query therefore hits it with probability $1/(N_0 - j)$. Padding to q_0 distinct queries can only help, and the probability of a hit among those queries is q_0/N_0.

Conditional on no hit, a final unqueried guess succeeds with probability at most $1/(N_0 - q_0)$; a queried guess cannot succeed. The overall probability is thus at most

$$\frac{q_0}{N_0} + \left(1 - \frac{q_0}{N_0}\right) \frac{1}{N_0 - q_0} = \frac{q_0 + 1}{N_0}.$$

Averaging over the private random tape proves the claim. □

Corollary G.3 (End-to-end oracle lower bound)**.** *Relativize either construction with $\mathcal{Q}_\Pi(x,u) = \mathbf{1}\{\Pi(u) = x\}$, for $n \ge 2$. The public code is independent of the sampled permutation: apart from the challenge x, access to Π is only through its forward oracle, with no permutation-dependent advice or uncharged preprocessing. Suppose compilation succeeds in producing an estimator of expected raw score at most $\eta < 1$ with probability at least p_{comp} on every challenge instance. If compilation followed by one fresh extraction makes at most q_0 forward queries in total, then*

$$q_0 \ge p_{\mathrm{comp}}(1 - \eta)2^n - 1.$$

For $p_{\mathrm{comp}} \ge 2/3$ and $\eta = 1/3$, this is $q_0 \ge (4/9)2^n - 1$, also at depth two and width n.

Proof. Conditional on successful compilation, the expected-payoff identity gives witness extraction probability at least $1 - \eta$. The combined procedure therefore inverts Π with probability at least $p_{\mathrm{comp}}(1 - \eta)$. Apply Lemma G.2; if $q_0 \ge 2^n$, the claimed lower bound is already automatic. □

Every forward query is charged, whether it occurs during compilation, gate evaluation, score simulation, returned-estimator evaluation, or candidate verification. The terminal oracle $\mathcal{O}$ remains constantly zero; Π is a separate forward oracle used by the public predicate. The bound concerns the complete compile-and-use procedure, not the compilation phase alone when later evaluation is unbounded.