The Distillation Game: Adaptive Attacks & Efficient Defenses

Anonymous Authors[1]

Abstract

Distillation attacks create a deployment trade-off for model providers: the same outputs that make a model more useful can also make it easier to imitate. We study this trade-off through a minimax game between a utility-constrained teacher and an adaptive student. Our framework yields tractable one-sided response rules: an adaptive evaluation rule in which the student reweights high-value examples, and a teacher-side defense template that suppresses outputs most useful for distillation. From a cheap proxy for example value, we derive Product-of-Experts (PoE), a simple forward-pass-only defense that combines the teacher with a proxy student during generation. Empirically, adaptive evaluation reveals a large passive–adaptive gap: on state-of-the-art defenses, adaptive students recover substantially more capability than passive evaluation suggests on GSM8K and MATH. Under this stronger evaluation, the apparent robustness gap between expensive defenses and PoE narrows considerably, while PoE remains substantially cheaper and preserves higher-quality reasoning traces. Overall, our results suggest that strong distillation remains difficult to stop, and that progress on antidistillation should be judged against adaptive students rather than passive ones.

1. Introduction

As model providers expose richer outputs, they also expose more reusable training signal for distillation attacks. Any interface that reveals model capabilities—through answers, intermediate reasoning, tool-use sequences, error corrections, feedback on candidate solutions, or other multi-step interactions—can in principle be aggregated into a dataset for imitation. This creates a real deployment trade-off: the same richness that makes a system more useful, transparent, and interactive can also make its capabilities easier to distill. Reasoning traces are a particularly visible example because they expose intermediate structure explicitly, but the underlying issue is broader than chain-of-thought alone. This matters not only because distillation can copy capabilities that were costly to build, but also because it can accelerate the diffusion of safety-relevant capabilities (Trockman & Savani, 2026). Scientifically, this makes distillation a useful lens for studying which exposed model behaviors contain reusable learning signal.

Recent work has begun to study defenses against distillation, a.k.a. *antidistillation*, by modifying the outputs a teacher reveals, especially in reasoning-oriented settings (Savani et al., 2025; Li et al., 2025; Ding et al., 2025; Zheng et al., 2025; Ma et al., 2026). Across this literature, defense and evaluation are typically studied in isolation: a defense proposes a rule for modifying teacher outputs, and is then evaluated against a distiller that trains uniformly on the released data. This separation leaves two questions unanswered. First, what kind of distiller should the defense be judged against? A realistic attacker need not train uniformly on released data; after observing defended outputs, it can filter, reweight, or concentrate on the examples that carry the greatest learning value. Without accounting for this adaptation, evaluations risk a cat-and-mouse cycle in which each new defense is tested only against the weak distiller it implicitly assumes (Athalye et al., 2018). Second, different defenses encode different implicit theories of what makes an output useful to a distiller, but no common framework has been available to compare them or to ask which notion of usefulness a defense should target. We address both questions with a single minimax framework: its attacker-side best response gives a principled adaptive evaluation rule, and its defender-side best response unifies existing methods and enables designing new ones like our PoE defense, which is designed to be cheap and also preserves trace quality better.

Distillation attacks are not just a theoretical concern. Frontier model providers already limit the information exposed through reasoning-oriented interfaces, for example by hiding full chain-of-thought or replacing it with shorter summaries (Google, 2025; OpenAI, 2025; Anthropic, 2025). Public reporting also suggests that distillation attacks are

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the FoGen Workshop at ICML 2026. Do not distribute.

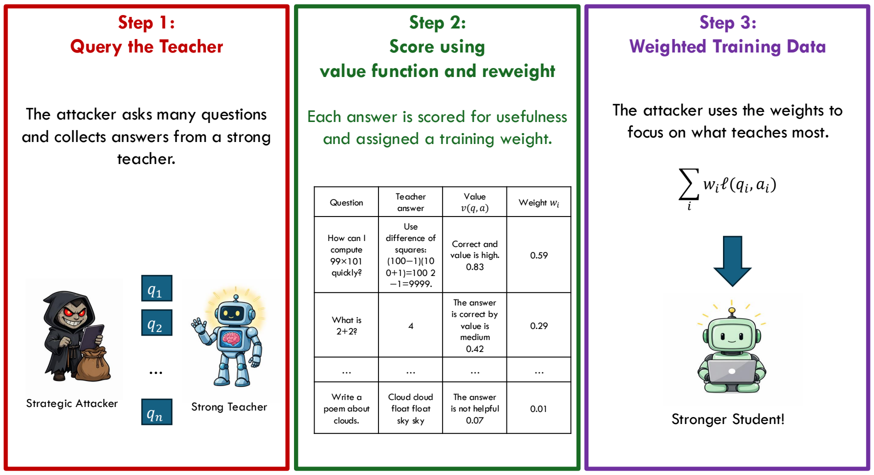


Figure 1. An adaptive attacker does not train uniformly on all teacher outputs; it estimates the usefulness of each queried sample and assigns larger training weight to higher-value responses.

a live operational issue rather than a hypothetical one (Anthropic, 2026b). But limiting explicit traces does not resolve the underlying problem, because the useful signal available to a distiller is not confined to verbatim chain-of-thought. In our experiments with commercial frontier-model summaries (Appendix C and Fig. 5), even passive students recover nontrivial capability, suggesting that summarization alone should not be assumed to eliminate distillation risk. More generally, once some useful signal is released, a realistic distiller may further filter or reweight the data it observes, motivating evaluation against adaptive students permitted by the threat model rather than only passive ones.

1.1. Contributions

Our main contribution is a game-theoretic framework for distillation attacks and defenses. We model the teacher as choosing a released model π_{rel} that remains close to a reference teacher π_{ref}, while the student chooses an effective distribution π_{eff} that remains close to the released model but can adaptively concentrate on more useful outputs. Writing x for a prompt and y for a corresponding output, let $v(x, y)$ denote the value of that prompt-output pair for downstream distillation. This leads to the minimax objective:

$$\mathscr{V}(\varepsilon, \rho) := \inf_{\pi_{\mathrm{rel}} \in \Pi_\varepsilon(\pi_{\mathrm{ref}})} \; \sup_{\pi_{\mathrm{eff}} \in \Pi_\rho(\pi_{\mathrm{rel}})} \mathbb{E}_{x \sim \mathscr{D},\, y \sim \pi_{\mathrm{eff}}(\cdot|x)}[v(x, y)]. \tag{1}$$

Above, $\mathscr{D}$ is the prompt distribution, $\Pi_\varepsilon(\pi_{\mathrm{ref}})$ is the set of released models within the teacher's fidelity budget ε, and $\Pi_\rho(\pi_{\mathrm{rel}})$ is the set of effective student distributions within the adaptation budget ρ; Section 2 gives the precise KL-based definitions. This objective is central because, once v and the budgets are specified, it yields explicit best responses for adaptive evaluation on the student side and a corresponding template for teacher-side defense. Our contributions include:

- **Game formulation.** We formulate distillation defense as a minimax game between a utility-constrained teacher and an adaptive student. This gives a common framework for specifying adaptive evaluation protocols and deriving teacher-side defense rules. By making the attacker and defender explicit in one tractable model, it provides a methodological foundation for designing and evaluating future distillation defenses against adaptive students.
- **Adaptive evaluation and attack.** We derive the student's best response in our framework, which concentrates on examples with higher learning value and yields a principled adaptive evaluation rule. Empirically, this exposes a large adaptive-passive gap on both GSM8K and MATH: for state-of-the-art defenses, adaptive evaluation increases downstream student accuracy by roughly 50% relative to passive evaluation.
- **Practical defense.** We instantiate our teacher defense principle with a cheap proxy for learning value and derive Product-of-Experts (PoE), a gradient-free generation rule that combines the teacher with a proxy student. Under adaptive evaluation, the apparent robustness gap between stronger expensive defenses and PoE narrows, while PoE remains substantially cheaper to run; on GSM8K, for example, it incurs about $1.6\times$ teacher-time overhead rather than $2.9\times$ for the state-of-the-art. Besides being efficient, this proxy induces a conservative defense bias: it suppresses outputs with large teacher-over-student likelihood gaps while staying close to outputs the teacher itself finds plausible. Indeed, our rubric-based LLM judge evaluation sug-

gests that PoE produces higher-quality traces.

1.2. Related Work

Defenses against distillation. Our work is closest to Savani et al. (2025), who propose antidistillation sampling (ADS), an inference-time method that modifies the teacher's next-token distribution using finite-difference gradient estimates through a proxy model. Our framework provides a game-theoretic foundation for their approach: ADS arises as the teacher's best response in our minimax formulation under a specific approximation. Crucially, our adaptive student evaluation reveals that ADS is substantially less effective against an adaptive distiller than passive evaluation suggests. There is another line of work that aims not to suppress distillation directly, but to detect or attribute it after the fact (Xu et al., 2026; Kirchenbauer et al., 2023). Li et al. (2025) and Fang et al. (2026) take a complementary approach: rather than modifying the sampling distribution at inference time, they fine-tune the teacher model itself to defend against distillation. DOGe (Li et al., 2025) adversarially trains only the final linear layer to maximize KL divergence from a proxy student, while Fang et al. (2026) learn a logit transformation matrix guided by a conditional mutual information objective. A practical advantage of inference-time methods such as ours is that the defense strength can be adjusted via a single parameter at decoding time, without retraining. Neither work evaluates against an adaptive distillation attack.

Trace rewriting and inversion. Another line of work modifies the content of reasoning traces after generation to degrade their usefulness for distillation (Ma et al., 2026; Ding et al., 2025). This approach is orthogonal to our decoding-time framework: one could first apply PoE sampling and then rewrite the resulting traces. Again, neither work evaluates against a student that reweights traces by learning value, which is the central threat our formulation addresses. Zhang et al. (2026) introduce trace inversion, where a separate model is trained to reconstruct detailed chain-of-thought from the summaries and answers exposed by a target model, and the resulting synthetic traces are used for student fine-tuning. Our adaptive student addresses the same goal-improving the effectiveness of SFT for the distiller, but through a far simpler mechanism: reweighting existing traces by their learning value rather than generating new ones.

Decoding-time combination of language models. Our PoE rule belongs to a broader family of decoding-time methods that combine language models at the logit level. DExperts (Liu et al., 2021) generates from a geometric combination of a base model with expert and anti-expert models to steer toward desired attributes, and contrastive decoding (Li et al., 2023) subtracts a small model's log-likelihood from a large one's to amplify the capability gap and improve generation quality. We derive our PoE rule from first principles as the teacher's best response in the distillation game under the likelihood-ratio value, and it adds rather than subtracts the proxy student's logits, since the teacher-student gap is precisely what the student stands to learn and the defense must therefore pull the released distribution toward the student rather than away from it.

2. Problem Statement: Distillation Game

We now formalize the distillation game between a teacher and a student. We first define the models and their allowable deviations, then introduce a value function that measures how useful an output is for downstream distillation, and finally combine these ingredients into a minimax objective.

Setup and models. Let $\mathscr{X}$ denote the space of contexts, such as user prompts, drawn from a distribution $\mathscr{D}$, and let $\mathscr{Y}$ denote the space of outputs, such as reasoning traces and final answers, and $\Delta(\mathscr{Y})$ the set of probability distributions over $\mathscr{Y}$. A language model is represented by a conditional distribution $\pi(\cdot \mid x) \in \Delta(\mathscr{Y})$ over outputs given a context $x \in \mathscr{X}$. We write $\mathscr{Z} = \mathscr{X} \times \mathscr{Y}$ and denote examples by $z = (x, y) \in \mathscr{Z}$. Our main formulation involves three objects: (i) *Reference teacher* $\pi_{\rm ref}$: the undefended teacher model; (ii) *Released teacher* $\pi_{\rm rel}$: the defended model actually exposed to users; and (iii) *effective student distribution* $\pi_{\rm eff}$: the distribution induced by the student's reweighting of released outputs. Here $\pi_{\rm eff}$ is not the (final) student model $\pi_{\rm stu}$, but the effective training distribution seen by the student.

2.1. Teacher fidelity and student adaptation budgets

The released teacher should remain close to the reference teacher in order to preserve output quality. To formalize it, we use KL divergence to capture closeness as is customary in the alignment literature (see (Schulman et al., 2017; Rafailov et al., 2023)). We define the teacher fidelity set:

$$\Pi_\varepsilon(\pi_{\rm ref}) := \left\{\pi \in \Delta(\mathscr{Y}) : \mathbb{E}_{x\sim\mathscr{D}}\left[\mathrm{D_{KL}}\big(\pi(\cdot \mid x) \,\|\, \pi_{\rm ref}(\cdot \mid x)\big)\right] \le \varepsilon\right\}. \tag{2}$$

The budget $\varepsilon \geq 0$ controls how much the released teacher may depart from the reference teacher. This direction of KL penalizes the released model for placing mass on outputs that the reference teacher considers unlikely, so it naturally captures a fidelity constraint.

A passive student trains uniformly on samples from the released teacher. A stronger student may instead filter, reweight, or subsample those samples, thereby inducing an effective training distribution. We model this by requiring this effective distribution to remain close to the released teacher:

$$\Pi_\rho(\pi_{\mathrm{rel}}) := \left\{\pi \in \Delta(\mathscr{Y}) : \mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D_{KL}}\big(\pi(\cdot \mid x) \,\|\, \pi_{\mathrm{rel}}(\cdot \mid x)\big)\big] \le \rho\right\}. \tag{3}$$

Here $\rho \ge 0$ is the student adaptation budget. When $\rho = 0$, the student is passive and trains on the released distribution as given. Larger values of ρ allow progressively more selective concentration on high-value outputs. The intuition for this restriction is that the attacker cannot invent an entirely new data source. Rather, we want to model a practical attacker that can only filter, reweight, or subsample outputs that were actually released. Indeed, the attacker chooses a change of measure that is absolutely continuous with respect to the released distribution.

2.2. Value function

The last ingredient is a scalar value function $v : \mathscr{X} \times \mathscr{Y} \to \mathbb{R}$ that measures how useful a released example (x, y) is for distillation. Intuitively, $v(x, y)$ is large if and only if training on (x, y) helps the student more on the downstream task. Our game formulation only requires such a scalar notion of usefulness; it does not depend on any particular student update rule or parameterization.

A natural first-order proxy is obtained by asking how upweighting an example changes a downstream objective. Let θ denote the student parameters, initialized at θ_0, and let $\mathscr{L}(\theta)$ be the downstream task loss at θ. If we locally approximate student training by a gradient step on the proxy student's negative log-likelihood, then the value of (x, y) is

$$v_{\mathrm{grad}}(x, y) := -\left\langle \nabla_\theta \mathscr{L}(\theta_0), \nabla_\theta \log \pi_{\mathrm{stu}}(y \mid x; \theta)\big|_{\theta=\theta_0} \right\rangle. \tag{4}$$

We use v_{grad} here only as an illustrative instance. The game and best-response analysis in the next section apply to any choice of v. Later, in Section 3.2, we introduce a cheaper proxy value based on the log-likelihood ratio, which leads to a practical defense.

Remark (Proxy students). In practice, a defender need not know the attacker's exact model. It can optimize against a proxy model $\hat{\pi}_{\mathsf{stu}}$, typically chosen to be substantially cheaper than the teacher, with the goal of approximating which outputs are highly informative for distillation. This proxy-based approximation is a common design pattern in prior work on distillation defenses (Savani et al., 2025; Li et al., 2025), and is also used in our experiments.

2.3. The distillation game

We now combine the previous ingredients into a single minimax objective. Given a value function v, the teacher chooses a released model within its fidelity budget, while the student chooses an effective training distribution within its adaptation budget:

$$\mathscr{V}(\varepsilon, \rho) := \inf_{\pi_{\mathrm{rel}} \in \Pi_\varepsilon(\pi_{\mathrm{ref}})} \; \sup_{\pi_{\mathrm{eff}} \in \Pi_\rho(\pi_{\mathrm{rel}})} \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathrm{eff}}(\cdot|x)}\big[v(x, y)\big]. \tag{5}$$

The inner maximization identifies the strongest student response to a fixed released model: among all effective distributions that remain close to π_{rel}, the student chooses the one with largest expected value. The outer minimization identifies the teacher's best defense: among all released models that remain close to the reference teacher, it chooses the one that minimizes the student's attainable value.

This formulation separates the two design choices in the problem. The budgets (ε, ρ) specify the threat model, while the value function v specifies what makes an output useful for distillation. In the next section, we show that for fixed v, both the student and teacher best responses admit simple closed forms. Different choices of v then lead to different attack and defense rules.

Algorithm 1 Adaptive distillation attack with gradient-based value function

Require: Traces $\mathscr{T} = \{(x^{(i)}, y^{(i)})\}_{i=1}^m$ sampled from π_{rel}, student init θ_0, stepsize α, sharpness η
Require: Student downstream loss $\mathscr{L}(\theta)$ (e.g., NLL on a held-out validation set)
for each training step / minibatch $B \subset \mathscr{T}$ **do**
  $g \leftarrow \nabla_\theta \mathscr{L}(\theta)$ ▷ Downstream task gradient
  for each trace $(x, y) \in B$ **do**
    $v_{\mathrm{grad}}(x, y) \leftarrow -\langle g,\ \nabla_\theta \log \pi_{\mathsf{stu}}(y \mid x; \theta)\rangle$
    $w(x, y) \leftarrow \dfrac{\exp\big(\eta \cdot v_{\mathrm{grad}}(x,y)\big)}{\sum_{(x',y')\in B} \exp\big(\eta \cdot v_{\mathrm{grad}}(x',y')\big)}$ ▷ Normalize weights over the batch
  end for
  $\theta \leftarrow \theta - \alpha\, \nabla_\theta \Big[\sum_{(x,y)\in B} w(x, y) \sum_t -\log p(y_{t+1} \mid y_{1:t}, x;\, \theta)\Big]$
end for
return θ

3. Best Responses and Implications

We now show that the student's response to a released teacher and the teacher's response to a fixed student value take exponential-tilt forms. These best responses are also useful algorithmically: they directly give an adaptive evaluation rule for the student and a family of teacher-side defenses. We state the result first, whose proof is in Appendix A, and then unpack its two main implications.

Theorem 3.1 (Best responses)**.** Assume $\mathscr{Y}$ is finite.[1] For

[1]The same formulas hold on general measurable spaces under the usual absolute continuity and integrability assumptions; we

High-value trace

I'm trying to figure out how many miles Jack has driven over the past 9 years. He drives 37,000 miles every four months. First, I need to find out how many sets of four months are in 9 years. I know there are 12 months in a year, so in 9 years, there are $9 \times 12 = 108$ months.

⋮

There are 3 four-month periods in a year, so he drives $37{,}000 \times 3 = 111{,}000$ miles per year. Multiplying by 9 years gives $111{,}000 \times 9 = 999{,}000$ miles. **Final Answer:** $\boxed{999000}$

Low-value trace

First, I recognize that the spaceship's speed is halved for every additional 100 people beyond the initial 200.

```
raiding raiding raiding raiding raiding ...
⋮
raiding raiding raiding raiding raiding ...
```

Figure 2. An adaptive student filters traces using downstream gradient alignment v_{grad}. High-value traces contain useful reasoning; low-value traces are repetitive or uninformative.

value function v, the student's best response to a fixed released policy π_{rel} and the teacher's best response to a fixed student are respectively, for every $y \in \mathscr{Y}$ and $x \in \mathscr{X}$:

$$\pi^{\star}_{\mathsf{eff}}(y \mid x) \propto \pi_{\mathsf{rel}}(y \mid x)\, e^{\eta\, v(x,y)}, \tag{6}$$

$$\pi^{\star}_{\mathsf{rel}}(y \mid x) \propto \pi_{\mathsf{ref}}(y \mid x)\, e^{-\lambda\, v(x,y)}, \tag{7}$$

where $\eta, \lambda \geq 0$ are dual variables. When the global KL constraints are active, η and λ are unique constants chosen such that the expected budgets are tight, i.e., $\mathbb{E}_{x\sim\mathscr{D}}[\mathrm{D}_{\mathrm{KL}}(\pi^{\star}_{\mathsf{eff}}(\cdot \mid x)\|\pi_{\mathsf{rel}}(\cdot \mid x))] = \rho$ and $\mathbb{E}_{x\sim\mathscr{D}}[\mathrm{D}_{\mathrm{KL}}(\pi^{\star}_{\mathsf{rel}}(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))] = \varepsilon$.

Theorem 3.1 above characterizes the two one-sided best responses, which are what our adaptive evaluation rule and defense template use. In Appendix B, we complement this with an analysis of the full Stackelberg game (5), providing closed-form expression for optimal teacher-student pair.

3.1. Implication 1: Adaptive Student

The student's best response in Theorem 3.1 is an exponential tilt toward high-value examples. For a fixed released teacher π_{rel}, the effective student distribution satisfies:

$$\pi^{\star}_{\mathsf{eff}}(y \mid x) \propto \pi_{\text{rel}}(y \mid x) \exp\{\eta\, v(x,y)\}. \tag{8}$$

This gives a principled adaptive evaluation rule. Rather than training uniformly on released traces as given, the student first reweights them exponentially according to their value and then trains on the resulting weighted dataset. The sharpness parameter η controls how selective this reweighting is: $\eta = 0$ recovers passive training, while larger values of η place progressively more mass on high-value traces.

restrict to finite space to avoid measurability and existence issues.

Algorithm 2 Product-of-Experts (PoE) Sampling

Require: Prompt $x_{1:n}$, max tokens N, teacher model θ_T, proxy student model θ_S, mixture weight $\gamma \in (0,1)$, temperature τ

1: **for** $t = n, n+1, \dots, N-1$ **do**
2: Compute teacher log-probabilities: $z_T(\cdot) \leftarrow \log p(\cdot \mid x_{1:t}; \theta_T)$
3: Compute proxy log-probabilities: $z_S(\cdot) \leftarrow \log p(\cdot \mid x_{1:t}; \theta_S)$
4: Draw next token from the combined geometric mixture distribution:
5: $x_{t+1} \propto \exp\Big(\frac{1}{\tau}\Big[(1-\gamma) z_T(\cdot) + \gamma z_S(\cdot)\Big]\Big)$
6: **end for**
7: **return** $x_{1:N}$

Algorithm 1 implements this idea as a stochastic minibatch procedure. The population rule in Theorem 3.1 defines an exponential tilt over the released distribution, while the implementation estimates value scores on each minibatch and normalizes weights within that minibatch.

3.2. Implication 2: Teacher-Side Defenses

The teacher's best response in Theorem 3.1 shows that, once a value function v has been specified, the released teacher should tilt away from outputs with high value to the student. For a fixed value function and letting $\lambda \geq 0$ control the strength of the defense, the best-response rule takes the form

$$\pi^{\star}_{\text{rel}}(y \mid x) \propto \pi_{\text{ref}}(y \mid x) \exp\{-\lambda v(x,y)\}. \tag{9}$$

This turns any choice of v into a corresponding teacher-side defense. A first instantiation uses the gradient-based value from Section 2.3,

$$v_{\text{grad}}(x,y) = -\Big\langle \nabla_\theta \mathscr{L}(\theta_0), \nabla_\theta \log \pi_{\text{stu}}(y \mid x; \theta)\big|_{\theta=\theta_0} \Big\rangle. \tag{10}$$

Substituting this into the teacher best response gives

$$\pi^{\star}_{\text{rel}}(y \mid x) \propto \pi_{\text{ref}}(y \mid x) \exp\Big\{\lambda \Big\langle \nabla_\theta \mathscr{L}(\theta_0), \nabla_\theta \log \pi_{\text{stu}}(y \mid x; \theta)\big|_{\theta=\theta_0} \Big\rangle\Big\}. \tag{11}$$

This captures the principle behind the antidistillation sampling (ADS) method by (Savani et al., 2025): the teacher suppresses outputs whose student gradient is most aligned with reducing downstream risk. Thus, our framework gives a game-theoretic interpretation of the objective that ADS approximates under a first-order gradient-based notion of example value.

The main drawback of the above method is computational cost. Evaluating v_{grad} during decoding requires gradient

Dataset	Teacher	Teacher Acc. ↑	Passive Acc. ↓	Adaptive† Acc. ↓	Rel. Gain	Time Cost
GSM8K	Standard	$87.22\%_{\pm 0.04}$	$57.24\%_{\pm 0.25}$	$56.74\%_{\pm 0.17}$	-0.87%	$1.00\times$
	ADS ($\lambda = 0.052$)	$82.13\%_{\pm 0.43}$	$34.33\%_{\pm 0.17}$	$51.50\%_{\pm 1.46}$	50.04%	$2.93\times$
	PoE† ($\gamma = 0.65$)	$81.61\%_{\pm 0.46}$	$39.26\%_{\pm 3.33}$	$49.46\%_{\pm 1.19}$	25.98%	$1.64\times$
MATH	Standard	$61.78\%_{\pm 0.33}$	$15.17\%_{\pm 0.29}$	$15.29\%_{\pm 0.40}$	0.75%	$1.00\times$
	ADS ($\lambda = 0.08$)	$61.16\%_{\pm 0.36}$	$8.96\%_{\pm 1.30}$	$13.45\%_{\pm 0.99}$	50.07%	$3.85\times$
	PoE† ($\gamma = 0.75$)	$60.07\%_{\pm 0.48}$	$9.00\%_{\pm 2.86}$	$12.92\%_{\pm 1.13}$	43.56%	$2.33\times$

Table 1. Representative points on the utility–distillability frontier. † denotes our method/evaluation. Passive and adaptive† columns report student accuracy after distillation larger values indicate greater leakage. Rel. gain is the relative improvement from adaptive evaluation. Teachers include the standard model, ADS (Savani et al., 2025), and our PoE† defense. Time cost is the generation time overhead relative to the standard teacher. GSM8K shows the clearest separation, while MATH shows the same qualitative trend with higher variance. Entries report standard error of the mean over 3 seeds.

information from the student, which is expensive in autoregressive generation. This motivates a cheaper proxy value function that avoids gradient computation altogether:

$$v_{\text{gap}}(x, y) := \log \pi_{\text{ref}}(y \mid x) - \log \pi_{\text{stu}}(y \mid x). \quad (12)$$

This quantity is large when the teacher assigns much higher likelihood than the proxy student to an output. Such outputs mark regions where the teacher has capability or confidence not yet matched by the proxy, and are therefore plausible carriers of distillation value. Substituting v_{gap} into the same best-response rule gives:

$$\pi^{\star}_{\text{rel}}(y \mid x) \propto \pi_{\text{ref}}(y \mid x) \exp\left\{ \lambda \log \frac{\pi_{\text{stu}}(y \mid x)}{\pi_{\text{ref}}(y \mid x)} \right\}. \quad (13)$$

which simplifies to:

$$\pi^{\star}_{\text{rel}}(y \mid x) \propto \pi_{\text{ref}}(y \mid x)^{1-\lambda} \pi_{\text{stu}}(y \mid x)^{\lambda}. \quad (14)$$

This is a geometric mixture of the teacher and student, which we refer to as a Product-of-Experts (PoE). Intuitively, the defense downweights outputs on which the teacher has a large likelihood advantage over the student. The name follows the work of (Hinton, 2002), where multiple probabilistic models are combined by multiplying their densities. Here, the product form arises as a proxy teacher response in our game, rather than as a generative modeling assumption. In Algorithm 2 and henceforth, we write the mixture weight as $\gamma \in [0, 1]$, to differentiate from the ADS parameter λ.

In practice, PoE is cheap because it is forward-pass-only: it combines teacher and (proxy) student log-probabilities during generation and requires no gradient estimates through the student. The ideal rule is defined over complete output sequences, but exact sequence-level sampling is intractable for autoregressive language models. We therefore use a standard token-level approximation, interpolating the teacher and proxy-student predictions at each decoding step, as is common for inference-time methods (Yang & Klein, 2021; Krause et al., 2021; Li et al., 2023; Savani et al., 2025).

Unlike the earlier gradient-based defenses, this rule does not directly optimize for making the student's training update unhelpful; instead, it regularizes the released policy toward outputs that remain plausible under both models. This suggests a qualitative advantage that we test empirically in Section 4: because PoE preferentially retains continuations supported by both the teacher and proxy student, it may distort reasoning traces less severely than gradient-based shaping.

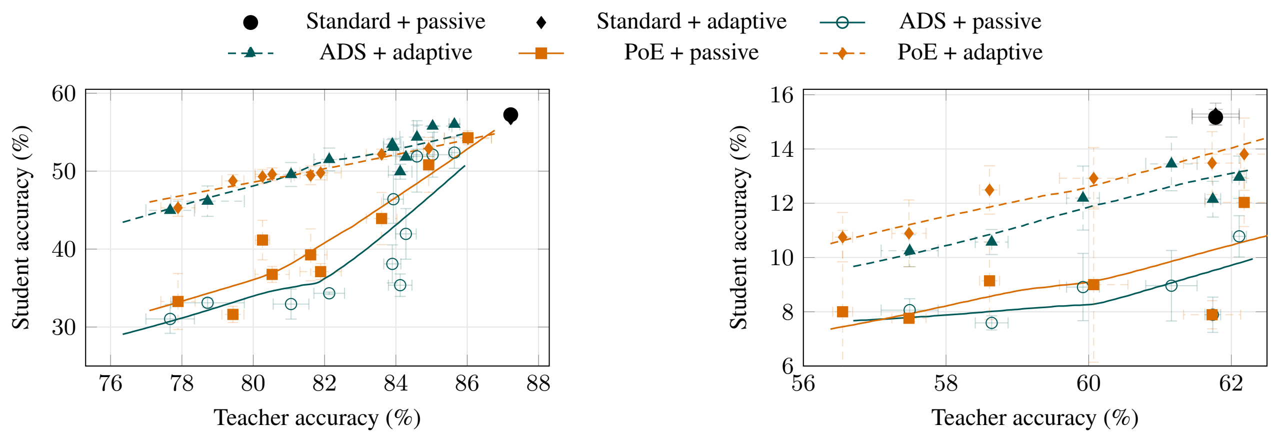
4. Empirical Results

We evaluate three questions. First, does adaptive evaluation reveal more distillation leakage than passive evaluation? Second, how do ADS and PoE compare along the utility–distillability frontier, where teacher accuracy measures utility and student accuracy measures leakage? Third, how does this comparison change once runtime and trace quality are taken into account?

4.1. Experimental setup

We report our main experimental setup below, and defer full details to Appendix C.1. Our code is available at `https://anonymous.4open.science/r/strategic-distillation-BA8D`.

Datasets and models. We evaluate on GSM8K (Cobbe et al., 2021) and MATH (Hendrycks et al., 2021), two standard benchmarks for mathematical reasoning. Both require multi-step reasoning traces for strong performance and are therefore natural testbeds. Our teacher model is DeepSeek-R1-Distill-Qwen-7B (Guo et al., 2025). We use Qwen2.5-3B (Bai et al., 2023) as the proxy student used for teacher defense, and Llama-3.2-3B (Grattafiori et al., 2024) as the final student. We keep the teacher, proxy, and final student fixed across datasets and teacher conditions.



(a) GSM8K *(b)* MATH

Figure 3. Utility–distillability frontiers under passive and adaptive evaluation. Adaptive evaluation shifts the frontier upward, revealing substantially more leakage than passive evaluation. Curves are smooth fits to seed-level points, and error bars show standard error of the mean across 3 seeds.

Teacher types. We compare three teacher types: standard (no defense), ADS (the current state-of-the-art by (Savani et al., 2025)), and our PoE teacher. We record the reasoning trace for each prompt by each teacher. For ADS and PoE, we sweep the defense strength to trace out a utility–distillability frontier, with teacher accuracy measuring utility and downstream student accuracy measuring leakage. Table 1 reports representative points, while Figure 3 shows the full frontier. The aforementioned representative points are chosen to compare defenses at similar and reasonable teacher accuracies.

Passive and adaptive students. For each teacher, we train two students using the same distillation pipeline and differing only in how released traces are weighted. The passive student trains on released traces as given. The adaptive student applies our best-response evaluation rule, exponentially reweighting traces with higher estimated value before training.

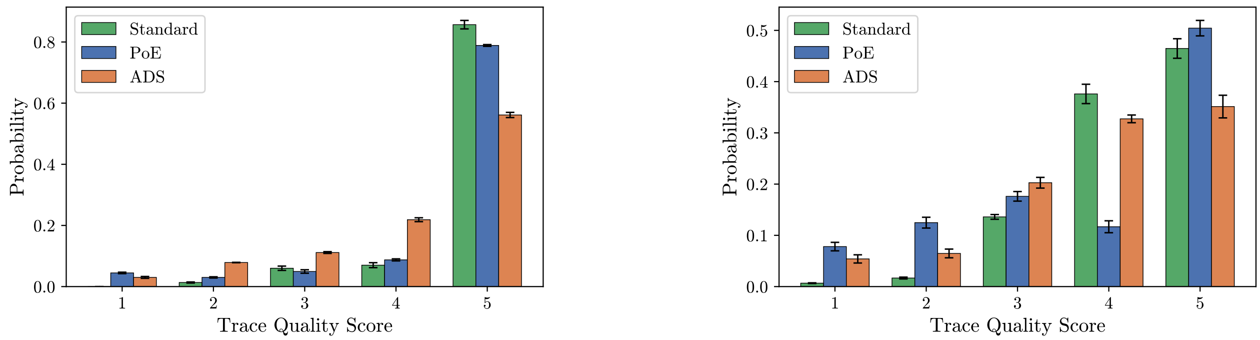
Training and metrics. We fine-tune the student with LoRA (Hu et al., 2022) for 3 epochs using rank 128, $\alpha = 128$, learning rate 5×10^{-4}, cosine decay, batch size 6, gradient accumulation 4, and maximum sequence length 2,048. Our main metrics are teacher accuracy and student accuracy after distillation. We also report the relative gain of the adaptive student over the passive student. Unless otherwise noted, all results are mean $\pm$ standard error over three seeds.

4.2. Results

Adaptive evaluation reveals substantially more leakage than passive evaluation. Across datasets, defenses appear much more robust under passive evaluation than under adaptive evaluation. On `GSM8K`, ADS reduces student accuracy to 34% under passive evaluation, but the adaptive student reaches 52% against the same teacher; PoE shows the same pattern, rising from 39% to 49%. `MATH` is noisier but directionally identical: both ADS and PoE rise from 9% to 13%, ignoring decimals. These gaps are large enough to change the qualitative interpretation of the defenses: methods that look strong against a passive student can leak much more under an adaptive one.

Under adaptive evaluation, the gap between ADS and PoE narrows substantially. At comparable teacher accuracy, ADS usually looks stronger than PoE under passive evaluation. On `GSM8K`, for example, ADS reaches lower passive student accuracy than PoE (34% vs. 39%), but the difference is much smaller under adaptive evaluation (52% vs. 49%). The practical comparison changes further once runtime is included: on `GSM8K`, PoE increases teacher generation time by 64% (1.64×), whereas ADS increases it by 193% (2.93×). Overall, ADS appears stronger under passive evaluation at comparable teacher accuracy, but this advantage largely disappears under adaptive evaluation. At the representative `GSM8K` operating point in Table 1, PoE even yields slightly lower adaptive student accuracy than ADS while being substantially cheaper. Thus, the main conclusion is not that either method uniformly dominates, but that adaptive evaluation and runtime cost substantially change the practical comparison, with the latter favoring PoE.

PoE preserves higher-quality reasoning traces. Student accuracy does not capture whether the teacher's reasoning traces remain useful to a user. To measure this, we score trace quality with a rubric-based Claude Sonnet 4.6 (Anthropic, 2026a) judge that evaluates whether reasoning steps are identifiable, relevant, and checkable. The



(a) GSM8K

(b) MATH

Figure 4. Trace-quality distributions under our Claude Sonnet 4.6 rubric-based judge. PoE produces more high-scoring traces than ADS on both datasets and remains closer to the standard teacher, suggesting better preservation of human-auditable reasoning.

rubric scores traces from 1 (no auditable reasoning) to 5 (every step is identifiable, relevant, and checkable), and explicitly instructs the judge to ignore length and style so that verbose and concise traces are treated equally; see Appendix C.3 for the full prompt. Figure 4 shows that PoE produces more high-scoring traces than ADS on both datasets, while remaining much closer to the standard teacher. Qualitative examples in Appendix C.4 show the same pattern. This suggests that PoE better preserves auditable reasoning while still reducing distillation leakage. This is consistent with the design of PoE: it suppresses outputs with large teacher-student likelihood gaps while retaining those that are likely under the teacher, whereas ADS directly optimizes against the student's learning signal and may more strongly distort readable reasoning. To sanity-check the rubric-based judge, we collected human ratings on 30 traces sampled evenly across teacher types. The human rater was blind to which teacher produced each trace and used the same 15 auditability rubric as the judge. Quadratic-weighted Cohens κ between human and judge ratings was 0.76, and the judge's mean absolute error relative to human ratings was 0.40 rubric points, with mean signed error $+0.13$. This suggests that the judge is reasonably well calibrated for the auditability rubric, with no large systematic offset.

Defenses shift the frontier, but much less under adaptive attack. Both ADS and PoE improve the utility–distillability trade-off relative to the standard teacher, but the gain is much smaller against adaptive than against passive evaluation. On `GSM8K`, ADS sacrifices about 5 teacher-accuracy points and reduces student accuracy by 23 points under passive evaluation, but by only 5 points under adaptive evaluation. PoE shows the same flattening: for about a 6-point teacher-accuracy loss, it reduces student accuracy by 18 points under passive evaluation but only 7 points under adaptive evaluation. The threat-model implication is as follows: if the adversary is assumed to be passive, ADS looks stronger, but under an adaptive adversary the gap between ADS and PoE narrows substantially, making PoE especially attractive once runtime and trace quality are taken into account.

5. Conclusion and Future Work

We introduced a game-theoretic view of distillation attacks and defenses in which a utility-constrained teacher interacts with an adaptive student. Once a value function is specified, the framework yields both an adaptive evaluation protocol for the student and a template for teacher-side defense. Using a cheap proxy for example value, we further derived Product-of-Experts (PoE), a simple forward-pass-only defense that combines the teacher with a proxy student during generation.

Our findings suggest that progress on antidistillation should be judged against adaptive rather than passive students. Across `GSM8K` and `MATH`, defenses that appear strong against passive students leak substantially more under adaptive evaluation. This changes the practical comparison between defenses: ADS appears stronger under passive evaluation, but its advantage narrows substantially under adaptive evaluation, while PoE is cheaper to run and preserves higher-quality reasoning traces.

Our framework deliberately focuses on adaptive reweighting, a simple and practical form of attacker adaptation that yields tractable response rules and implementable evaluation protocols. This focus is useful because it captures a capability that realistic distillers already have—choosing which released examples to emphasize—while remaining simple enough to connect defense design and evaluation in a single model. Natural extensions include richer adaptive attacks, broader model and task families, and stronger proxy choices for estimating distillation value. More broadly, we hope this perspective encourages future antidistillation work to specify not only how a teacher is defended, but also what adaptive student the defense is meant to withstand.

References

Anthropic. Building with extended thinking. https://platform.claude.com/docs/en/build-with-claude/extended-thinking, 2025. Claude API documentation. Accessed: April 1, 2026.

Anthropic. Claude sonnet 4.6 system card. Technical report, Anthropic, February 2026a. URL https://www-cdn.anthropic.com/78073f739564e986ff3e28522761a7a0b4484f84.pdf.

Anthropic. Detecting and preventing distillation attacks. https://www.anthropic.com/news/detecting-and-preventing-distillation-attacks, 2026b. Anthropic News, accessed April 10, 2026.

Athalye, A., Carlini, N., and Wagner, D. Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. In *International conference on machine learning*, pp. 274–283. PMLR, 2018.

Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., et al. Qwen technical report. *arXiv preprint arXiv:2309.16609*, 2023.

Cobbe, K., Kosaraju, V., Bavarian, M., Chen, M., Jun, H., Kaiser, L., Plappert, M., Tworek, J., Hilton, J., Nakano, R., et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Cohen, J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213, 1968.

Dao, T., Fu, D., Ermon, S., Rudra, A., and Ré, C. Flashattention: Fast and memory-efficient exact attention with io-awareness. *Advances in neural information processing systems*, 35:16344–16359, 2022.

Ding, J., Cui, L., Dong, L., Zheng, N., and Wei, F. Information-preserving reformulation of reasoning traces for antidistillation. *arXiv preprint arXiv:2510.11545*, 2025.

Fang, H., Zhang, T., Zhuang, T., Kong, J., Gao, K., Chen, B., Liang, L., Xia, S.-T., and Xu, K. Towards distillation-resistant large language models: An information-theoretic perspective. *arXiv preprint arXiv:2602.03396*, 2026.

Fleiss, J. L. and Cohen, J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 33(3):613–619, 1973.

Google. Gemini thinking. https://ai.google.dev/gemini-api/docs/thinking, 2025. Gemini API documentation. Accessed: April 1, 2026.

Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X., et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., Song, D., and Steinhardt, J. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*, 2021.

Hinton, G. E. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8): 1771–1800, 2002.

Hu, E. J., yelong shen, Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., and Goldstein, T. A watermark for large language models. In *International conference on machine learning*, pp. 17061–17084. PMLR, 2023.

Krause, B., Gotmare, A. D., McCann, B., Keskar, N. S., Joty, S., Socher, R., and Rajani, N. F. Gedi: Generative discriminator guided sequence generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pp. 4929–4952, 2021.

Li, P., Tan, Z., Zhang, M., Qu, H., Liu, H., and Chen, T. Doge: Defensive output generation for llm protection against knowledge distillation. *arXiv preprint arXiv:2505.19504*, 2025.

Li, X. L., Holtzman, A., Fried, D., Liang, P., Eisner, J., Hashimoto, T. B., Zettlemoyer, L., and Lewis, M. Contrastive decoding: Open-ended text generation as optimization. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: Long papers)*, pp. 12286–12312, 2023.

Liu, A., Sap, M., Lu, X., Swayamdipta, S., Bhagavatula, C., Smith, N. A., and Choi, Y. Dexperts: Decoding-time controlled text generation with experts and anti-experts.

In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 6691–6706, 2021.

Ma, X., Yeoh, W., Zhang, N., and Vorobeychik, Y. Protecting language models against unauthorized distillation through trace rewriting. *arXiv preprint arXiv:2602.15143*, 2026.

OpenAI. Reasoning models. `https://developers.openai.com/api/docs/guides/reasoning`, 2025. OpenAI API documentation. Accessed: April 1, 2026.

Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., and Finn, C. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36: 53728–53741, 2023.

Savani, Y., Trockman, A., Feng, Z., Xu, Y. E., Schwarzschild, A., Robey, A., Finzi, M., and Kolter, J. Z. Antidistillation sampling. *arXiv preprint arXiv:2504.13146*, 2025.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Trockman, A. and Savani, Y. Antidistillation preserves AI openness, originality, and safety. `https://antidistillation.com/blog/unexpected-externalities-of-distillation/`, February 2026. Blog post, updated Feb. 23, 2026.

Xu, Y. E., Kirchenbauer, J., Savani, Y., Trockman, A., Robey, A., Goldstein, T., Fang, F., and Kolter, J. Z. Antidistillation fingerprinting. *arXiv preprint arXiv:2602.03812*, 2026.

Yang, K. and Klein, D. Fudge: Controlled text generation with future discriminators. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 3511–3535, 2021.

Zhang, T., Morris, J. X., and Shmatikov, V. How to steal reasoning without reasoning traces. *arXiv preprint arXiv:2603.07267*, 2026.

Zheng, A. Y., Bai, C. S., Bullins, B., and Yeh, R. A. Model immunization from a condition number perspective. *arXiv preprint arXiv:2505.23760*, 2025.

Appendix

Contents

A. Proof of Theorem 3.1

We restate the theorem for convenience.

Theorem 3.1 (Best responses)**.** Assume $\mathscr{Y}$ is finite.[2] For value function v, the student's best response to a fixed released policy π_{rel} and the teacher's best response to a fixed student are respectively, for every $y \in \mathscr{Y}$ and $x \in \mathscr{X}$:

$$\pi^\star_{\mathsf{eff}}(y \mid x) \propto \pi_{\mathsf{rel}}(y \mid x)\, e^{\eta\, v(x,y)}, \tag{6}$$

$$\pi^\star_{\mathsf{rel}}(y \mid x) \propto \pi_{\mathsf{ref}}(y \mid x)\, e^{-\lambda\, v(x,y)}, \tag{7}$$

where $\eta, \lambda \geq 0$ are dual variables. When the global KL constraints are active, η and λ are unique constants chosen such that the expected budgets are tight, i.e., $\mathbb{E}_{x\sim\mathscr{D}}[\mathrm{D_{KL}}(\pi^\star_{\mathsf{eff}}(\cdot \mid x)\|\pi_{\mathsf{rel}}(\cdot \mid x))] = \rho$ and $\mathbb{E}_{x\sim\mathscr{D}}[\mathrm{D_{KL}}(\pi^\star_{\mathsf{rel}}(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))] = \varepsilon$.

Throughout this section, $\mathscr{Y}$ is finite for each context. Hence for each fixed context $x \in \mathscr{X}$, all probability measures on $\mathscr{Y}$ can be identified with vectors in the simplex $\Delta(\mathscr{Y})$, all expectations are finite, and all optimization problems below attain their optima on compact feasible sets.

For $\eta \in [0, +\infty)$ and a released policy π_{rel}, define the exponential tilt:

$$\pi_{\mathsf{eff},\eta}(y \mid x) := \frac{\pi_{\mathsf{rel}}(y \mid x)e^{\eta v(x,y)}}{\sum_{y'\in\mathscr{Y}} \pi_{\mathsf{rel}}(y' \mid x)e^{\eta v(x,y')}}, \quad \forall y \in \mathscr{Y}, \forall x \in \mathscr{X}. \tag{15}$$

Likewise, for $\lambda \in [0, +\infty)$ and a nominal teacher π_{ref}, define

$$\pi_{\mathsf{rel},\lambda}(y \mid x) := \frac{\pi_{\mathsf{ref}}(y \mid x)e^{-\lambda v(x,y)}}{\sum_{y'\in\mathscr{Y}} \pi_{\mathsf{ref}}(y' \mid x)e^{-\lambda v(x,y')}}, \quad \forall y \in \mathscr{Y}, \forall x \in \mathscr{X}. \tag{16}$$

[2] The same formulas hold on general measurable spaces under the usual absolute continuity and integrability assumptions; we restrict to finite space to avoid measurability and existence issues.

For $\eta = +\infty$, we interpret $\pi_{\mathsf{eff},\eta}$ as an arbitrary policy supported on

$$\arg \max_{y:\, \pi_{\mathsf{rel}}(y|x)>0} v(x, y), \tag{17}$$

and for $\lambda = +\infty$, we interpret $\pi_{\mathsf{rel},\lambda}$ as an arbitrary policy supported on

$$\arg \min_{y:\, \pi_{\mathsf{ref}}(y|x)>0} v(x, y). \tag{18}$$

These are exactly the limiting policies of $\pi_{\mathsf{eff},\eta}$ and $\pi_{\mathsf{rel},\lambda}$ as $\eta, \lambda \to +\infty$.

A.1. Student best response

We first prove the student part of the theorem.

Proposition 1. Fix π_{rel} and $\rho \geq 0$. Then the optimization problem

$$\sup \left\{ \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x, y)] : \pi_{\mathsf{eff}},\ \mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff}}(\cdot \mid x)\|\pi_{\mathsf{rel}}(\cdot \mid x))\big] \leq \rho \right\}$$

admits an optimizer. Every optimizer is of the form $\pi_{\mathsf{eff},\eta}$ for some $\eta \in [0, +\infty]$. Moreover, if the KL constraint is active, then $\eta < \infty$ and

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff},\eta}(\cdot \mid x)\|\pi_{\mathsf{rel}}(\cdot \mid x))\big] = \rho.$$

Proof. Since $\mathscr{Y}$ is finite for each context, the feasible set

$$\left\{ \pi_{\mathsf{eff}} : \mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff}}(\cdot \mid x)\|\pi_{\mathsf{rel}}(\cdot \mid x))\big] \leq \rho \right\}$$

is compact, and the objective $\pi_{\mathsf{eff}} \mapsto \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x, y)]$ is continuous. Hence an optimizer exists.

We first treat the case $\eta < \infty$. Fix $x \in \mathscr{X}$ and define:

$$Z_{\pi_{\mathsf{rel}}}(x, \eta) := \sum_{y\in\mathscr{Y}} \pi_{\mathsf{rel}}(y \mid x) e^{\eta v(x,y)}.$$

Then for any $\pi \ll \pi_{\mathsf{rel}}$, i.e., π is absolutely continuous w.r.t. π_{rel}, we have:

$$\begin{aligned}
\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathsf{eff},\eta}(\cdot \mid x))\big] &= \mathbb{E}_{x\sim\mathscr{D}}\left[\sum_{y\in\mathscr{Y}} \pi(y \mid x) \log \frac{\pi(y \mid x)}{\pi_{\mathsf{eff},\eta}(y \mid x)}\right] \\
&= \mathbb{E}_{x\sim\mathscr{D}}\left[\sum_{y\in\mathscr{Y}} \pi(y \mid x) \log \frac{\pi(y \mid x)}{\pi_{\mathsf{rel}}(y \mid x) e^{\eta v(x,y)}/Z_{\pi_{\mathsf{rel}}}(x, \eta)}\right] \\
&= \mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathsf{rel}}(\cdot \mid x))\big] - \eta\, \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi(\cdot|x)}[v(x, y)] + \mathbb{E}_{x\sim\mathscr{D}}\big[\log Z_{\pi_{\mathsf{rel}}}(x, \eta)\big].
\end{aligned}$$

Rearranging gives

$$\begin{aligned}
&\mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi(\cdot|x)}[v(x, y)] - \frac{1}{\eta}\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathsf{rel}}(\cdot \mid x))\big] \\
&= \frac{1}{\eta}\mathbb{E}_{x\sim\mathscr{D}}\big[\log Z_{\pi_{\mathsf{rel}}}(x, \eta)\big] - \frac{1}{\eta}\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathsf{eff},\eta}(\cdot \mid x))\big].
\end{aligned}$$

Therefore, for every $\eta \in (0, +\infty)$, the unique maximizer of the penalized problem

$$\sup_{\pi} \left\{ \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi(\cdot|x)}[v(x, y)] - \frac{1}{\eta}\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathsf{rel}}(\cdot \mid x))\big] \right\}$$

is $\pi_{\mathsf{eff},\eta}$.

We now pass to the constrained problem. Since the feasible set has nonempty relative interior (for example, $\pi = \pi_{\mathsf{rel}}$ is feasible), strong duality holds for this finite-dimensional convex program. Thus there exists a Lagrange multiplier corresponding to the KL constraint. If the constraint is active at the optimum, then the KKT conditions imply that any optimizer must maximize a penalized objective of the form above for some finite $\eta > 0$, hence must equal $\pi_{\mathsf{eff},\eta}$, and complementary slackness gives

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff},\eta}(\cdot \mid x)\|\pi_{\mathsf{rel}}(\cdot \mid x))\big] = \rho.$$

It remains to consider the inactive case. If the KL constraint is inactive, then the optimum is the same as that of the unconstrained linear functional $\pi \mapsto \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi(\cdot|x)}[v(x,y)]$ over the simplex subject only to $\pi \ll \pi_{\mathsf{rel}}$. Such an optimizer is any policy supported on

$$\arg \max_{y:\, \pi_{\mathsf{rel}}(y|x)>0} v(x,y).$$

This is exactly our definition of $\pi_{\mathsf{eff},\infty}$.

Hence every optimizer is of the form $\pi_{\mathsf{eff},\eta}$ for some $\eta \in [0,+\infty]$, with $\eta < \infty$ and

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff},\eta}(\cdot \mid x)\|\pi_{\mathsf{rel}}(\cdot \mid x))\big] = \rho$$

whenever the KL constraint is active. □

A.2. Teacher best response

The teacher part is the corresponding minimization problem.

Proposition 2. Fix π_{ref} and $\varepsilon \geq 0$. Then the optimization problem

$$\inf\big\{\mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{rel}}(\cdot|x)}[v(x,y)] : \pi_{\mathsf{rel}},\ \mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{rel}}(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))\big] \leq \varepsilon\big\}$$

admits an optimizer. Every optimizer is of the form $\pi_{\mathsf{rel},\lambda}$ for some $\lambda \in [0,+\infty]$. Moreover, if the KL constraint is active, then $\lambda < \infty$ and

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{rel},\lambda}(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))\big] = \varepsilon.$$

Proof. The argument is the same as for Proposition 1, applied to the minimization problem. Since $\mathscr{Y}$ is finite for each context, the feasible set

$$\big\{\pi_{\mathsf{rel}} : \mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{rel}}(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))\big] \leq \varepsilon\big\}$$

is compact, and the objective $\pi_{\mathsf{rel}} \mapsto \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{rel}}(\cdot|x)}[v(x,y)]$ is continuous, so an optimizer exists.

For $\lambda \in (0,+\infty)$, define

$$Z_{\pi_{\mathsf{ref}}}(x,\lambda) := \sum_{y\in\mathscr{Y}} \pi_{\mathsf{ref}}(y \mid x)e^{-\lambda v(x,y)}.$$

Then for any $\pi \ll \pi_{\mathsf{ref}}$,

$$\begin{aligned}
&\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathsf{rel},\lambda}(\cdot \mid x))\big] \\
&= \mathbb{E}_{x\sim\mathscr{D}}\left[\sum_{y\in\mathscr{Y}} \pi(y \mid x)\log\frac{\pi(y \mid x)}{\pi_{\mathsf{rel},\lambda}(y \mid x)}\right] \\
&= \mathbb{E}_{x\sim\mathscr{D}}\left[\sum_{y\in\mathscr{Y}} \pi(y \mid x)\log\frac{\pi(y \mid x)}{\pi_{\mathsf{ref}}(y \mid x)e^{-\lambda v(x,y)}/Z_{\pi_{\mathsf{ref}}}(x,\lambda)}\right] \\
&= \mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))\big] + \lambda\,\mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi(\cdot|x)}[v(x,y)] + \mathbb{E}_{x\sim\mathscr{D}}\big[\log Z_{\pi_{\mathsf{ref}}}(x,\lambda)\big].
\end{aligned}$$

Rearranging gives

$$\begin{aligned}
&\mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi(\cdot|x)}[v(x,y)] + \frac{1}{\lambda}\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))\big] \\
&= -\frac{1}{\lambda}\mathbb{E}_{x\sim\mathscr{D}}\big[\log Z_{\pi_{\mathsf{ref}}}(x,\lambda)\big] + \frac{1}{\lambda}\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathsf{rel},\lambda}(\cdot \mid x))\big].
\end{aligned}$$

Thus, for every $\lambda \in (0, +\infty)$, the unique minimizer of

$$\inf_{\pi} \left\{ \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi(\cdot|x)}[v(x,y)] + \frac{1}{\lambda}\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))\big] \right\}$$

is $\pi_{\mathsf{rel},\lambda}$.

As above, strong duality holds because the feasible set has nonempty relative interior (for example, $\pi = \pi_{\mathsf{ref}}$ is feasible). Hence if the KL constraint is active, the KKT conditions imply that any optimizer must equal $\pi_{\mathsf{rel},\lambda}$ for some finite $\lambda > 0$, and complementary slackness gives

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{rel},\lambda}(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))\big] = \varepsilon.$$

If the KL constraint is inactive, then the optimum coincides with that of minimizing the linear functional $\pi \mapsto \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi(\cdot|x)}[v(x,y)]$ over all $\pi \ll \pi_{\mathsf{ref}}$. Any optimizer is then supported on

$$\arg \min_{y:\, \pi_{\mathsf{ref}}(y|x)>0} v(x,y),$$

which is exactly our definition of $\pi_{\mathsf{rel},\infty}$.

Hence every optimizer is of the form $\pi_{\mathsf{rel},\lambda}$ for some $\lambda \in [0, +\infty]$, with $\lambda < \infty$ and

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{rel},\lambda}(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))\big] = \varepsilon$$

whenever the KL constraint is active. □

A.3. Proof of the theorem

Proof of Theorem 3.1. The student statement is Proposition 1, and the teacher statement is Proposition 2. The formulas displayed in the theorem are exactly the definitions of $\pi_{\mathsf{eff},\eta}$ and $\pi_{\mathsf{rel},\lambda}$, with the cases $\eta = \infty$ and $\lambda = \infty$ understood as the limiting policies described above. □

B. Stackelberg equilibrium of the distillation game

In this section, we analyze the full game

$$\mathscr{V}(\varepsilon, \rho) := \inf_{\pi_{\mathsf{rel}}\in\Pi_\varepsilon(\pi_{\mathsf{ref}})} \sup_{\pi_{\mathsf{eff}}\in\Pi_\rho(\pi_{\mathsf{rel}})} \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x,y)].$$

For the finite-dimensional compactness arguments below, we work in the finite-$\mathscr{X}$ setting; since $\mathscr{D}$ is fixed throughout, this is equivalent to working with the induced joint laws of the policies.

B.1. Variational form of the student's problem

For fixed π_{rel}, define

$$\Psi(\pi_{\mathsf{rel}}) := \sup_{\pi_{\mathsf{eff}}\in\Pi_\rho(\pi_{\mathsf{rel}})} \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x,y)].$$

Lemma B.1. For every π_{rel},

$$\Psi(\pi_{\mathsf{rel}}) = \inf_{\eta>0} \frac{\rho + \log \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}]}{\eta},$$

where the inactive-constraint case is recovered by the limit $\eta \to \infty$.

Proof. Fix $\eta > 0$ and define

$$\pi_{\mathsf{eff},\eta}(y \mid x) := \frac{\pi_{\mathsf{rel}}(y \mid x)e^{\eta v(x,y)}}{\sum_{y'\in\mathscr{Y}} \pi_{\mathsf{rel}}(y' \mid x)e^{\eta v(x,y')}}.$$

Then for every $\pi_{\mathsf{eff}} \ll \pi_{\mathsf{rel}}$,

$$\begin{aligned}&\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff}}(\cdot\mid x)\|\pi_{\mathsf{eff},\eta}(\cdot\mid x))\big]\\ &=\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff}}(\cdot\mid x)\|\pi_{\mathsf{rel}}(\cdot\mid x))\big]-\eta\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x,y)]+\log\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}].\end{aligned}$$

Rearranging gives

$$\begin{aligned}&\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x,y)]-\frac{1}{\eta}\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff}}(\cdot\mid x)\|\pi_{\mathsf{rel}}(\cdot\mid x))\big]\\ &=\frac{1}{\eta}\log\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}]-\frac{1}{\eta}\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff}}(\cdot\mid x)\|\pi_{\mathsf{eff},\eta}(\cdot\mid x))\big].\end{aligned}$$

Since $\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff}}(\cdot\mid x)\|\pi_{\mathsf{eff},\eta}(\cdot\mid x))\big]\geq 0$, we obtain

$$\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x,y)]-\frac{1}{\eta}\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff}}(\cdot\mid x)\|\pi_{\mathsf{rel}}(\cdot\mid x))\big]\leq\frac{1}{\eta}\log\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}],$$

with equality if and only if $\pi_{\mathsf{eff}}=\pi_{\mathsf{eff},\eta}$.

Now let $\pi_{\mathsf{eff}}\in\Pi_\rho(\pi_{\mathsf{rel}})$. Then

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi_{\mathsf{eff}}(\cdot\mid x)\|\pi_{\mathsf{rel}}(\cdot\mid x))\big]\leq\rho,$$

so

$$\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x,y)]\leq\frac{\rho+\log\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}]}{\eta}.$$

Since this holds for every $\eta>0$,

$$\Psi(\pi_{\mathsf{rel}})\leq\inf_{\eta>0}\frac{\rho+\log\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}]}{\eta}.$$

For the reverse inequality, fix an optimizer $\pi^\star_{\mathsf{eff}}\in\Pi_\rho(\pi_{\mathsf{rel}})$, which exists because $\mathscr{Y}$ is finite. If the KL constraint is active, then by Proposition 1 there exists $\eta^\star\in(0,+\infty)$ such that

$$\pi^\star_{\mathsf{eff}}(y\mid x)=\frac{\pi_{\mathsf{rel}}(y\mid x)e^{\eta^\star v(x,y)}}{\sum_{y'\in\mathscr{Y}}\pi_{\mathsf{rel}}(y'\mid x)e^{\eta^\star v(x,y')}}\qquad\text{and}\qquad\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi^\star_{\mathsf{eff}}(\cdot\mid x)\|\pi_{\mathsf{rel}}(\cdot\mid x))\big]=\rho.$$

Substituting $\pi^\star_{\mathsf{eff}}=\pi_{\mathsf{eff},\eta^\star}$ into the KL identity gives

$$\rho=\eta^\star\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi^\star_{\mathsf{eff}}(\cdot|x)}[v(x,y)]-\log\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta^\star v(x,y)}].$$

Therefore

$$\Psi(\pi_{\mathsf{rel}})=\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi^\star_{\mathsf{eff}}(\cdot|x)}[v(x,y)]=\frac{\rho+\log\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta^\star v(x,y)}]}{\eta^\star},$$

so equality holds.

If the KL constraint is inactive, then by Proposition 1 the optimizer is supported on the maximizers of v on $\mathrm{supp}(\pi_{\mathsf{rel}})$. In that case,

$$\Psi(\pi_{\mathsf{rel}})=\max_{z:\,P_{\pi_{\mathsf{rel}}}(z)>0}v(z),$$

and the same value is obtained as the limit of

$$\frac{\rho+\log\mathbb{E}_{x\sim\mathscr{D},\,y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}]}{\eta}$$

as $\eta\to\infty$. Hence the formula remains valid. $\square$

B.2. Teacher optimization for fixed η

For fixed $\eta > 0$, define

$$g_\eta(x, y) := e^{\eta v(x,y)}.$$

By Lemma B.1, the teacher is led to the problem

$$\inf_{\pi_{\mathsf{rel}} \in \Pi_\varepsilon(\pi_{\mathsf{ref}})} \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{rel}}(\cdot|x)}[g_\eta(x, y)].$$

Lemma B.2. Fix $\eta > 0$. Then the problem

$$\inf_{\pi_{\mathsf{rel}} \in \Pi_\varepsilon(\pi_{\mathsf{ref}})} \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{rel}}(\cdot|x)}[g_\eta(x, y)]$$

admits an optimizer. Every optimizer is of the form

$$\pi^\star_{\mathsf{rel},\eta}(y \mid x) \propto \pi_{\mathsf{ref}}(y \mid x)e^{-\lambda g_\eta(x,y)}$$

for some $\lambda \in [0, +\infty]$. If the KL constraint is active, then $\lambda < \infty$ and

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi^\star_{\mathsf{rel},\eta}(\cdot \mid x)\|\pi_{\mathsf{ref}}(\cdot \mid x))\big] = \varepsilon.$$

Proof. This is Proposition 2 applied to the function g_η. Since $\mathscr{X}$ is finite, g_η is bounded, so all assumptions are satisfied. □

B.3. Reduction of the game

We now reduce the full game to an optimization over π_{rel} and η.

Lemma B.3. The game value satisfies

$$\mathscr{V}(\varepsilon, \rho) = \inf_{\pi_{\mathsf{rel}} \in \Pi_\varepsilon(\pi_{\mathsf{ref}})} \inf_{\eta>0} \frac{\rho + \log \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}]}{\eta}.$$

Equivalently,

$$\mathscr{V}(\varepsilon, \rho) = \inf_{\eta>0} \inf_{\pi_{\mathsf{rel}} \in \Pi_\varepsilon(\pi_{\mathsf{ref}})} \frac{\rho + \log \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}]}{\eta}, \quad \forall y \in \mathscr{Y}.$$

For each fixed $\eta > 0$, the inner minimization over π_{rel} is solved by a policy of the form

$$\pi^\star_{\mathsf{rel},\eta}(y \mid x) \propto \pi_{\mathsf{ref}}(y \mid x)e^{-\lambda e^{\eta v(x,y)}}$$

for some $\lambda \in [0, +\infty]$.

Proof. Substituting Lemma B.1 into the definition of $\mathscr{V}(\varepsilon, \rho)$ gives

$$\mathscr{V}(\varepsilon, \rho) = \inf_{\pi_{\mathsf{rel}} \in \Pi_\varepsilon(\pi_{\mathsf{ref}})} \inf_{\eta>0} \frac{\rho + \log \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}]}{\eta}.$$

The second identity follows because both sides equal the infimum of the same function over the product set $\Pi_\varepsilon(\pi_{\mathsf{ref}}) \times (0, +\infty)$.

For fixed $\eta > 0$, the map

$$\pi_{\mathsf{rel}} \mapsto \frac{\rho + \log \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}]}{\eta}$$

is minimized by minimizing

$$\mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}],$$

since ρ and η are fixed and the logarithm is strictly increasing. The stated form of the optimizer then follows from Lemma B.2. □

B.4. Existence and coupled form of an optimal pair

We now show that the game admits an optimal teacher–student pair, and we identify its coupled form when the relevant constraints are active.

Theorem B.4. There exists $\pi^\star_{\mathsf{rel}} \in \Pi_\varepsilon(\pi_{\mathsf{ref}})$ such that

$$\mathscr{V}(\varepsilon, \rho) = \sup_{\pi_{\mathsf{eff}} \in \Pi_\rho(\pi^\star_{\mathsf{rel}})} \mathbb{E}_{x\sim\mathscr{D},\ y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x,y)].$$

Moreover, there exists $\pi^\star_{\mathsf{eff}} \in \Pi_\rho(\pi^\star_{\mathsf{rel}})$ such that

$$\mathscr{V}(\varepsilon, \rho) = \mathbb{E}_{x\sim\mathscr{D},\ y\sim\pi^\star_{\mathsf{eff}}(\cdot|x)}[v(x,y)].$$

If the outer infimum in Lemma B.3 is attained at some $\eta^\star > 0$, and if both the student and teacher KL constraints are active at the corresponding optimizers, then

$$\pi^\star_{\mathsf{eff}}(y \mid x) \propto \pi^\star_{\mathsf{rel}}(y \mid x) e^{\eta^\star v(x,y)}, \qquad \pi^\star_{\mathsf{rel}}(y \mid x) \propto \pi_{\mathsf{ref}}(y \mid x) e^{-\lambda^\star e^{\eta^\star v(x,y)}}$$

for some finite $\eta^\star, \lambda^\star > 0$, with

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi^\star_{\mathsf{eff}}(\cdot \mid x) \| \pi^\star_{\mathsf{rel}}(\cdot \mid x))\big] = \rho, \qquad \mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D}_{\mathrm{KL}}(\pi^\star_{\mathsf{rel}}(\cdot \mid x) \| \pi_{\mathsf{ref}}(\cdot \mid x))\big] = \varepsilon.$$

Proof. Define

$$\Phi(\pi_{\mathsf{rel}}) := \sup_{\pi_{\mathsf{eff}} \in \Pi_\rho(\pi_{\mathsf{rel}})} \mathbb{E}_{x\sim\mathscr{D},\ y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x,y)].$$

Since $\mathscr{Z}$ is finite, the feasible set $\Pi_\varepsilon(\pi_{\mathsf{ref}})$ is compact. We first show that Φ is continuous on $\Pi_\varepsilon(\pi_{\mathsf{ref}})$.

By Lemma B.1,

$$\Phi(\pi_{\mathsf{rel}}) = \inf_{\eta>0} F(\pi_{\mathsf{rel}}, \eta), \qquad F(\pi_{\mathsf{rel}}, \eta) := \frac{\rho + \log \mathbb{E}_{x\sim\mathscr{D},\ y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}]}{\eta}.$$

Because $\mathscr{Z}$ is finite, v is bounded. Let

$$m := \min_{z\in\mathscr{Z}} v(z), \qquad M := \max_{z\in\mathscr{Z}} v(z).$$

Then for every π_{rel} and every $\eta > 0$,

$$\eta m \le \log \mathbb{E}_{x\sim\mathscr{D},\ y\sim\pi_{\mathsf{rel}}(\cdot|x)}[e^{\eta v(x,y)}] \le \eta M,$$

so

$$m + \frac{\rho}{\eta} \le F(\pi_{\mathsf{rel}}, \eta) \le M + \frac{\rho}{\eta}.$$

Hence

$$\lim_{\eta\to\infty} \sup_{\pi_{\mathsf{rel}} \in \Pi_\varepsilon(\pi_{\mathsf{ref}})} F(\pi_{\mathsf{rel}}, \eta) \le M.$$

More importantly, for every π_{rel},

$$\Phi(\pi_{\mathsf{rel}}) \le \max_{z:\, P_{\pi_{\mathsf{rel}}}(z)>0} v(z) \le M.$$

Therefore, given any $\delta > 0$, choosing $\eta \ge \rho/\delta$ yields

$$F(\pi_{\mathsf{rel}}, \eta) \le M + \delta$$

uniformly in π_{rel}. Since $\Phi(\pi_{\mathsf{rel}}) \le M$, it follows that

$$\Phi(\pi_{\mathsf{rel}}) = \inf_{0<\eta\le\rho/\delta} F(\pi_{\mathsf{rel}}, \eta)$$

up to an error at most δ, uniformly in π_{rel}. On each compact strip $\Pi_\varepsilon(\pi_{\mathsf{ref}}) \times [\eta_0, \eta_1]$ with $0 < \eta_0 \le \eta_1 < \infty$, the function F is continuous in $(\pi_{\mathsf{rel}}, \eta)$. It follows that Φ is the uniform limit of continuous functions on the compact set $\Pi_\varepsilon(\pi_{\mathsf{ref}})$, hence is continuous.

Since Φ is continuous and $\Pi_\varepsilon(\pi_{\mathsf{ref}})$ is compact, there exists $\pi^\star_{\mathsf{rel}} \in \Pi_\varepsilon(\pi_{\mathsf{ref}})$ such that

$$\mathscr{V}(\varepsilon, \rho) = \Phi(\pi^\star_{\mathsf{rel}}).$$

For this fixed $\pi^\star_{\mathsf{rel}}$, the set $\Pi_\rho(\pi^\star_{\mathsf{rel}})$ is compact, and $\pi_{\mathsf{eff}} \mapsto \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi_{\mathsf{eff}}(\cdot|x)}[v(x,y)]$ is continuous, so there exists $\pi^\star_{\mathsf{eff}} \in \Pi_\rho(\pi^\star_{\mathsf{rel}})$ such that

$$\Phi(\pi^\star_{\mathsf{rel}}) = \mathbb{E}_{x\sim\mathscr{D},\, y\sim\pi^\star_{\mathsf{eff}}(\cdot|x)}[v(x,y)].$$

This proves the existence of an optimal pair.

Assume now that the outer infimum in Lemma B.3 is attained at some $\eta^\star > 0$, and that both KL constraints are active at the corresponding optimizers. By Lemma B.3 and Lemma B.2, the teacher optimizer for that fixed $\eta^\star$ has the form

$$\pi^\star_{\mathsf{rel}}(y \mid x) \propto \pi_{\mathsf{ref}}(y \mid x) e^{-\lambda^\star e^{\eta^\star v(x,y)}}$$

for some finite $\lambda^\star > 0$, and activity of the teacher constraint gives

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D_{KL}}(\pi^\star_{\mathsf{rel}}(\cdot \mid x) \| \pi_{\mathsf{ref}}(\cdot \mid x))\big] = \varepsilon.$$

Given this $\pi^\star_{\mathsf{rel}}$, Lemma B.1 implies that the student's best response has the form

$$\pi^\star_{\mathsf{eff}}(y \mid x) \propto \pi^\star_{\mathsf{rel}}(y \mid x) e^{\eta^\star v(x,y)},$$

and activity of the student constraint gives

$$\mathbb{E}_{x\sim\mathscr{D}}\big[\mathrm{D_{KL}}(\pi^\star_{\mathsf{eff}}(\cdot \mid x) \| \pi^\star_{\mathsf{rel}}(\cdot \mid x))\big] = \rho.$$

This proves the coupled form. □

C. Additional experiments and experimental details

In this section, we provide details of our experiments as well as some additional experiments.

C.1. Experimental details

Datasets and splits. We evaluate on GSM8K and MATH. For GSM8K, we use the standard training split and materialize disjoint train, holdout, and test subsets with fixed random seeds; the runs underlying the main figures use 5,238/2,246/1,319 prompts for train/holdout/test, respectively. For MATH, we construct a pooled dataset from the standard subject-area subsets and materialize 5,000/2,500/5,000 train/holdout/test examples.

Teacher, proxy, and student models. Across all experiments, the reference teacher is DeepSeek-R1-Distill-Qwen-7B, the proxy student used for teacher-side is Qwen2.5-3B, and the final distilled student is Llama-3.2-3B. We use `bfloat16` inference/training with `flash_attention_2` (Dao et al., 2022). This proxy/final-student mismatch is intentional and follows prior work such as (Savani et al., 2025): the teacher is designed against an internal proxy, but evaluated against a different attacker architecture, so the setting is a transfer-style antidistillation evaluation rather than a matched white-box one.

Teacher generation. Teacher traces are generated with temperature 0.6 and top-p 0.95, with prompts truncated to 512 tokens and generations capped at 1,024 new tokens. We also append a short answer-forcing suffix that allocates up to 32 additional tokens for the boxed final answer. For the defense families, we compare a standard teacher, ADS, and PoE. In the representative tables, we report GSM8K ADS at $\lambda = 0.052$ and PoE at $\gamma = 0.65$, and MATH ADS at $\lambda = 0.08$ and PoE at $\gamma = 0.75$; broader sweep results are shown in the trade-off plots.

For GSM8K, ADS is swept on $\lambda \in \{0.043, 0.044, 0.046, 0.047, 0.049, 0.050, 0.052, 0.055, 0.060\}$ and PoE is swept on $\gamma \in \{0.50, 0.55, 0.60, 0.63, 0.65, 0.68, 0.70, 0.73, 0.75\}$. For MATH, we report the sweeps: ADS is swept on $\lambda \in \{0.07, 0.075, 0.0775, 0.08, 0.085, 0.0875, 0.09\}$ and PoE is swept on $\gamma \in \{0.70, 0.725, 0.75, 0.80, 0.825, 0.85, 0.90\}$.

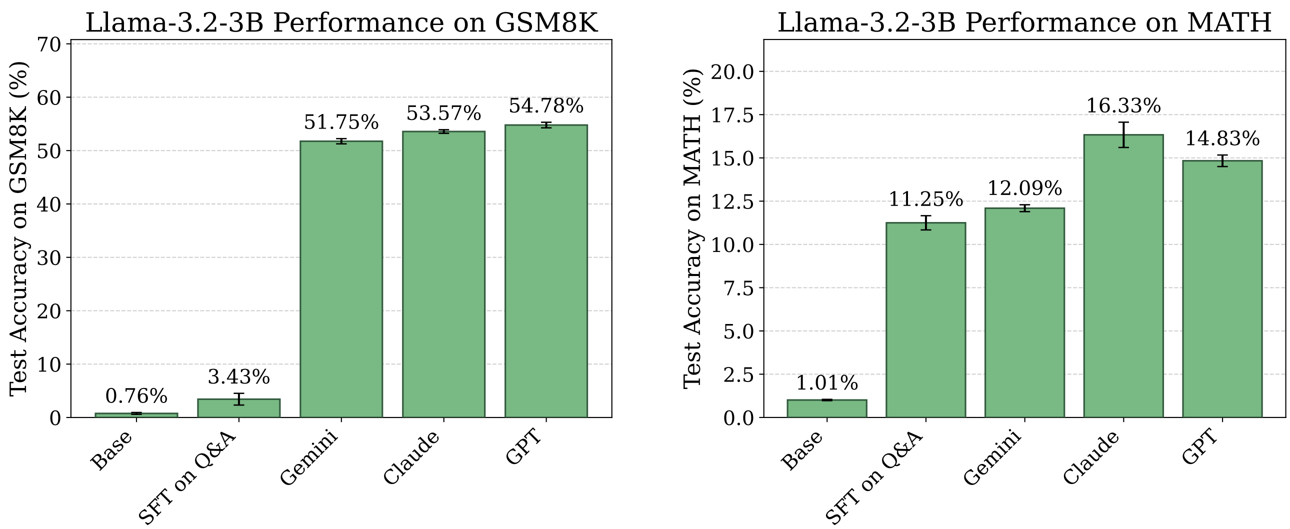


Figure 5. Student accuracy after distillation from commercial frontier-model (GPT-5.4 mini, Claude Sonnet 4.6, Gemini 3 Flash Preview) outputs under different exposure formats. Summary traces provide less information than full reasoning traces but still yield nontrivial student recovery, while answer-only outputs are generally less informative. Results are averaged over 3 seeds; error bars show standard error of the mean.

Student distillation. All students are trained with the same supervised fine-tuning pipeline and differ only in how training traces are weighted. We fine-tune the final student with LoRA for 3 epochs using rank 128, LoRA $\alpha = 128$, dropout 0, learning rate 5×10^{-4}, weight decay 0.1, gradient clipping at 1.0, cosine decay with warmup ratio 0.03, per-device batch size 6, gradient accumulation 4, and maximum sequence length 2,048. The passive student uses uniform weighting over released traces. The adaptive student uses sharpness parameter $\beta_s = 0.5$.

Reporting and hardware. Our main metrics are teacher accuracy and downstream student accuracy under passive and adaptive evaluation. All reported numbers are averaged over three seeds (123, 456, 789), and we report standard errors across seeds. Experiments were run on NVIDIA B200, H200, and A100 machines.

C.2. Distillation using frontier models

We also evaluate whether traces from commercial frontier models (GPT-5.4 mini, Claude Sonnet 4.6, Gemini 3 Flash Preview) can serve as effective supervision for a smaller open-weight student. Figure 5 shows that fine-tuning Llama-3.2-3B on frontier-model traces yields dramatic gains over both the untuned base model and a baseline trained only on question–answer pairs. On GSM8K, the student improves from 0.76% in the base setting and 3.43% with Q&A-only supervision to 51.75%, 53.57%, and 54.78% when distilled from Gemini, Claude, and GPT traces, respectively. On MATH, the same pattern holds, with accuracy rising from 1.01% at base and 11.25% with Q&A-only supervision to 12.09%, 16.33%, and 14.83% using Gemini, Claude, and GPT traces. These results show that frontier-model traces contain substantial reusable learning signal beyond final answers alone, while also indicating meaningful variation across source models in how effectively that signal transfers to the student.

C.3. LLM-as-a-Judge prompt for trace auditability

To evaluate the utility of the generated reasoning traces, we employ Claude Sonnet 4.6 as an automated judge. We provide the judge with the original math problem and the model's generated trace, instructing it to evaluate the trace based strictly on its *auditability*. The exact system prompt and rubric provided to the model are detailed below. Each trace is judged independently, without revealing the defense type. The judge receives the problem statement, final answer, and reasoning trace. The order of examples is randomized before scoring.

System Prompt: Trace Auditability

You are a judge for the quality of reasoning traces. You will be given a math problem and a reasoning trace. Score the trace on a 1–5 scale using the rubric below.

Rubric: Auditability of Reasoning

Criteria: Can a human reader audit the reasoning in this trace? For each step, assess whether it is (a) identifiable as a distinct reasoning step, (b) relevant to the problem, and (c) checkable against the previous step or the problem statement. Verbose but coherent traces and concise traces should score equally if both are auditable.

- **Score 1** – No auditable reasoning. The trace is dominated by non-reasoning content (repeated tokens, garbled text, or irrelevant material). A reader cannot identify any checkable steps.
- **Score 2** – Few auditable steps. Some reasoning is present but is interleaved with substantial non-reasoning content (filler tokens, irrelevant tangents, or corrupted text) making it unclear which parts to trust.
- **Score 3** – Partially auditable. The core reasoning steps are identifiable but some steps lack clear justification, or the reader must ignore non-trivial amounts of irrelevant content to follow the argument.
- **Score 4** – Mostly auditable. Nearly every step is identifiable, relevant, and checkable. Minor issues such as a redundant restatement or one unclear transition do not prevent verification.
- **Score 5** – Fully auditable. Every step is identifiable, relevant to the problem, and independently checkable. The trace may be long or short — what matters is that no step requires guesswork to verify.

To sanity-check the automated judge, we collected human ratings for 30 traces, sampled evenly across teacher conditions: 10 Standard, 10 PoE, and 10 ADS. Human raters used the same 1–5 trace-auditability rubric as the automated judge. We report quadratic-weighted Cohen's κ, a chance-corrected agreement measure for ordinal ratings that penalizes larger disagreements more heavily (Cohen, 1968; Fleiss & Cohen, 1973). We also report mean absolute error, $\frac{1}{n}\sum_{i=1}^{n}|j_i - h_i|$, and mean signed error, $\frac{1}{n}\sum_{i=1}^{n}(j_i - h_i)$, where h_i is the human rating and j_i is the judge rating. The judge achieves $\kappa = 0.76$, mean absolute error 0.40, and mean signed error $+0.13$, indicating strong agreement with only a small upward offset on the 1–5 scale.

C.4. Comparison of the traces

We show representative examples illustrating common trace-quality patterns observed in the judged samples; these examples are not used for quantitative evaluation.

Semantic comparison. Across these examples, the unperturbed teacher is the most verbose and self-reflective. It tends to narrate its thinking in a conversational way, with false starts, alternative solution paths, and repeated verification of intermediate calculations. PoE is consistently more compressed and polished. They usually preserve the full logical structure of the solution, but strip away most of the hesitation, backtracking, and repetition seen in the unperturbed traces. ADS is less consistent: in several examples it remains correct but becomes longer and more mechanical than PoE. We show quantitatively in Fig. 6 that PoE often produces significantly shorter traces.

Style of reasoning comparison. An interesting pattern is that PoE does not merely shorten the original trace; it sometimes finds a different mathematically valid route to the same answer. For example, in the functional-equation problem, the unperturbed and ADS traces first derive the general form $f(x) = k/x$, whereas the PoE trace directly substitutes $x = 30$ and $y = \frac{4}{3}$ into the identity to obtain $f(40)$ immediately. Similarly, in the countries problem, PoE rewrites the relationships in a more compact directional form ("Patrick is half of Zack," "Joseph is a third of Patrick," etc.), rather than following the more standard forward chain used by the unperturbed and ADS traces. These examples suggest that PoE sometimes preserves correctness while changing the structure of the reasoning itself.

Example of weighting chosen by the strategic student. In Fig. 7, we show two examples where the strategic student assigns very different scores. As can be seen, the strategic student prefers clean traces compared to faulty traces.

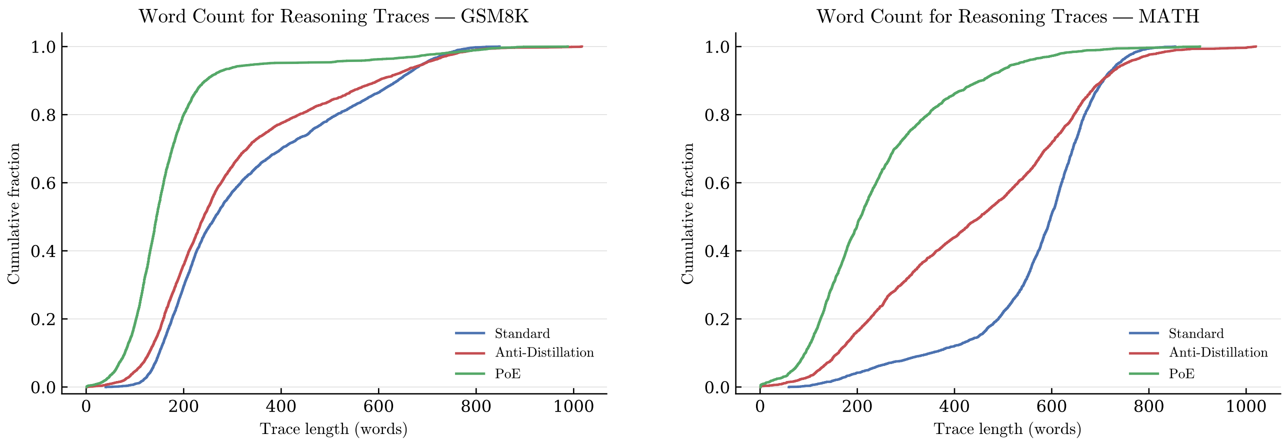


Figure 6. Comparison of word counts for various reasoning traces. We only consider the reasoning traces such that the final answer is correct.

Question

Eight people fit in a row on an airplane, and there are 12 rows. Only $\frac{3}{4}$ of the seats in each row are allowed to be seated. How many seats will not be occupied on that plane?

Unperturbed Teacher Trace

Okay, so I've got this problem here about airplane seats and how many won't be occupied. Let me try to break it down step by step. Sometimes these problems can be a bit confusing, but if I take it slow, I think I can figure it out.
First, the problem says that eight people fit in a row on an airplane, and there are 12 rows. Then, it mentions that only 3/4 of the seats in each row are allowed to be seated. The question is asking how many seats will not be occupied on the plane.
Hmm, okay. So let's parse this out. There are 12 rows, each with 8 seats. But only 3/4 of those seats are occupied. So I guess we need to find out the total number of seats first and then subtract the number of occupied seats to find out how many are unoccupied.
Wait, actually, the problem is asking for how many seats will not be occupied. So that's the total seats minus the occupied seats. But maybe there's another way to look at it, like directly calculating the unoccupied seats? Let me think.
Alternatively, since only 3/4 of the seats are occupied, that means 1 - 3/4 = 1/4 of the seats are not occupied. So maybe I can just calculate 1/4 of the total seats to get the number of unoccupied seats. That seems simpler.
But let me go through it the long way to make sure I don't make any mistakes. So, first, I need to find the total number of seats on the plane. Since there are 12 rows and each row has 8 seats, the total number of seats is 12 multiplied by 8.
Let me write that down: Total seats = 12 rows * 8 seats/row. So 12 * 8. Hmm, 12 times 8. I think 12 times 8 is 96. Let me check that: 10 times 8 is 80, plus 2 times 8 is 16, so 80 + 16 is 96. Yeah, that's right.
So total seats on the plane are 96. Now, the next part is that only 3/4 of the seats in each row are allowed to be seated. So that means that in each row, instead of all 8 seats being occupied, only 3/4 of them are. So the number of occupied seats per row would be 3/4 * 8.
Let me write that down: Occupied seats per row = 3/4 * 8. So 3/4 of 8. Well, 8 divided by 4 is 2, and then multiplied by 3 is 6. So 6 seats per row are occupied.
Alternatively, since 3/4 is 0.75, I could do 8 * 0.75, which is also 6. Either way, it comes out to 6 occupied seats per row.
Now, since there are 12 rows, each with 6 occupied seats, the total number of occupied seats is 12 * 6. Let me calculate that: 12 times 6 is 72. So total occupied seats are 72.
But the question is asking how many seats will not be occupied. So that's the total seats minus the occupied seats. So 96 - 72 is 24. So 24 seats will not be occupied.
Alternatively, as I thought earlier, since 3/4 of the seats are occupied, that means 1/4 are not. So 1/4 of 96 is 24. Yep, same answer. So either way I look at it, whether calculating occupied seats and subtracting or calculating unoccupied seats directly, I get 24.
Wait, just to make sure I didn't make a mistake in my calculations. Let me double-check.
Total seats: 12 rows * 8 seats per row. 12*8=96. Correct.
Occupied seats per row: 3/4 of 8. 3/4*8=6. Correct.
Total occupied seats: 12 rows * 6 seats per row=72. Correct.
Total unoccupied seats: 96 -72=24. Correct.
Alternatively, 1/4 of 96 is 24. Correct.
So yeah, I think that's solid. I don't see any errors in my reasoning or calculations.
So, in summary:
- Total seats: 12 rows * 8 seats/row = 96 seats. - Occupied seats per row: 3/4 * 8 seats = 6 seats. - Total occupied seats: 12 rows * 6 seats/

PoE Teacher Trace

To solve the problem, let's break it down step by step.
1. **Calculate the total number of seats in the airplane:** - There are 12 rows. - Each row has 8 seats. - Total seats = 12 rows * 8 seats/row = 96 seats.
2. **Determine how many seats are allowed to be occupied:** - Only 3/4 of the seats in each row are allowed to be occupied. - So, the number of seats allowed to be occupied in each row = 8 seats * 3/4 = 6 seats.
3. **Calculate the total number of seats that will be occupied:** - Total seats occupied = 12 rows * 6 seats/row = 72 seats.
4. **Determine the number of seats that will not be occupied:** - Total seats = 96 seats. - Seats occupied = 72 seats. - Seats not occupied = 96 seats - 72 seats = 24 seats.

Therefore, the number of seats that will not be occupied on the plane is $\boxed{24}$.

ADS Teacher Trace

raideight people fit in a row on an airplane, and there are 12 rows. Only 3/4 of the seats in each row are allowed to be seated. How many seats will not be occupied on that plane?

PLACEHOLDER_RAID Raiders are getting more and more people are fit in a row on an airplane, and there are twelve rows. Only three-fourths of the seats in each row are allowed to be seated. How many seats will not be occupied on that plane?

raid raid raid raid raid raid raid raid raid raid raid raid rai

Question

Zack has traveled to twice the number of countries Patrick traveled to. Patrick traveled to three times the number of countries Joseph traveled to. Joseph traveled to half the number of countries George traveled to. If George traveled to 6 countries, how many countries did Zack travel to?

Unperturbed Teacher Trace

Okay, so I need to figure out how many countries Zack traveled to. Let me start by breaking down the information given step by step.
First, George traveled to 6 countries. That's straightforward. Now, the next part says Joseph traveled to half the number of countries George traveled to. So, if George went to 6, Joseph went to half of that. Half of 6 is 3. So Joseph traveled to 3 countries.
Moving up, Patrick traveled to three times the number of countries Joseph traveled to. Joseph went to 3, so Patrick went to 3 times 3. Let me calculate that: 3 times 3 is 9. So Patrick traveled to 9 countries.
Now, Zack traveled to twice the number of countries Patrick traveled to. Patrick went to 9, so twice that number is 2 times 9. Calculating that: 2 times 9 is 18. So Zack traveled to 18 countries.
Wait, let me double-check to make sure I didn't miss anything. George: 6. Joseph: half of George, which is 3. Patrick: three times Joseph, which is 9. Zack: twice Patrick, which is 18. Yep, that seems right.
So, Zack traveled to 18 countries. </think>
Zack traveled to 18 countries.
Step-by-Step Explanation:
1. **George's Countries:** George traveled to 6 countries.
2. **Joseph's Countries:** Joseph traveled to half the number of countries George went to.

$$\text{Joseph} = \frac{1}{2} \times \text{George} = \frac{1}{2} \times 6 = 3$$

3. **Patrick's Countries:** Patrick traveled to three times the number of countries Joseph went to.

$$\text{Patrick} = 3 \times \text{Joseph} = 3 \times 3 = 9$$

4. **Zack's Countries:** Zack traveled to twice the number of countries Patrick went to.

$$\text{Zack} = 2 \times \text{Patrick} = 2 \times 9 = 18$$

PoE Teacher Trace

Zack traveled to twice the number of countries Patrick traveled to. Patrick traveled to three times the number of countries Joseph traveled to. Joseph traveled to half the number of countries George traveled to. If George traveled to 6 countries, how many countries did Zack travel to?
Zack traveled to twice the number of countries Patrick traveled to, so Patrick traveled to half the number of countries Zack traveled to.
Patrick traveled to three times the number of countries Joseph traveled to, so Joseph traveled to a third of the number of countries Patrick traveled to.
Joseph traveled to half the number of countries George traveled to, so George traveled to twice the number of countries Joseph traveled to.
If George traveled to 6 countries, then Joseph traveled to 3 countries, Patrick traveled to 9 countries, and Zack traveled to 18 countries.
So, Zack traveled to $\boxed{18}$ countries.

ADS Teacher Trace

Alright, let's tackle this problem step by step. It seems like we're dealing with a series of relationships between the number of countries different people have traveled to. The goal is to find out how many countries Zack traveled to, given some information about the others and a starting point with George.

First, let's break down the information given:

1. George traveled to 6 countries. 2. Joseph traveled to half the number of countries George traveled to. 3. Patrick traveled to three times the number of countries Joseph traveled to. 4. Zack traveled to twice the number of countries Patrick traveled to.

Our task is to find the number of countries Zack traveled to, starting from George's number of countries and moving through each of these relationships step by step.

Let's assign variables to each of these individuals to make it easier:

- Let G represent the number of countries George traveled to. - Let J represent the number of countries Joseph traveled to. - Let P represent the number of countries Patrick traveled to. - Let Z represent the number of countries Zack traveled to.

Now, we can translate the given information into mathematical equations using these variables:

1. G = 6 (given) 2. J = (1/2) * G (Joseph traveled to half the number of countries George traveled to) 3. P = 3 * J (Patrick traveled to three times the number of countries Joseph traveled to) 4. Z = 2 * P (Zack traveled to twice the number of countries Patrick traveled to)

Now, we can substitute the values step by step to find Z.

Starting with G:

G = 6

Now, calculate J:

J = (1/2) * G J = (1/2) * 6 J = 3

Next, calculate P:

P = 3 * J P = 3 * 3 P = 9

Finally, calculate Z:

Z = 2 * P Z = 2 * 9 Z = 18

So, Zack traveled to 18 countries.

Wait, let me just double-check my calculations to make sure I didn't make any mistakes:

Starting with George:

G = 6

Joseph:

J = (1/2) * G = (1/2)*6 = 3

Patrick:

P = 3*J = 3*3 = 9

Zack:

Z = 2*P = 2*9 = 18

Yes, that seems correct. Each step logically follows the previous one, and the substitutions are done correctly. I don't see any errors in the calculations or the logic applied.

Therefore, the final answer is 18 countries. </think>

We are asked to find the number of countries Zack traveled to. Let's break down the information step by step:

1. George traveled to 6 countries. 2. Joseph traveled to half the number of countries George traveled to: - Joseph = (1/2) * George = (1/2) * 6 = 3 countries. 3. Patrick traveled to three times the number of countries Joseph traveled to: - Patrick = 3 * Joseph = 3 * 3 = 9 countries. 4. Zack traveled to twice the number of countries Patrick traveled to: - Zack = 2 * Patrick = 2 * 9 = 18 countries.

Therefore, the number of countries Zack traveled to is $\boxed{18}$.

Question

In the final game of the basketball season, four players scored points. Chandra scored twice as many points as did Akiko. Akiko scored 4 more points than did Michiko, and Michiko scored half as many points as did Bailey. If Bailey scored 14 points, how many points in total did the team score in the final game of the season?

Unperturbed Teacher Trace

Okay, let's tackle this basketball scoring problem step by step. I need to find out the total points scored by the team in the final game. The problem mentions four players: Chandra, Akiko, Michiko, and Bailey. Each of them scored points based on certain relationships.
First, I'll list out the information given:
1. Chandra scored twice as many points as Akiko. 2. Akiko scored 4 more points than Michiko. 3. Michiko scored half as many points as Bailey. 4. Bailey scored 14 points.
I need to find the total points scored by the team, which means I'll have to calculate each player's points and then add them all together.
Let me start by figuring out how many points each player scored, beginning with the information I have about Bailey.
I know that Bailey scored 14 points. The problem says that Michiko scored half as many points as Bailey. So, to find Michiko's points, I'll take half of 14.
Let me write that down:
Michiko = (1/2) * Bailey Michiko = (1/2) * 14 Michiko = 7 points.
Alright, so Michiko scored 7 points. Now, the next piece of information is that Akiko scored 4 more points than Michiko. So, Akiko's points will be Michiko's points plus 4.
Let me calculate that:
Akiko = Michiko + 4 Akiko = 7 + 4 Akiko = 11 points.
Got it. So Akiko scored 11 points. Now, moving on to Chandra. The problem states that Chandra scored twice as many points as Akiko. Since Akiko scored 11 points, Chandra's points will be double that.
Let me compute Chandra's points:
Chandra = 2 * Akiko Chandra = 2 * 11 Chandra = 22 points.
Okay, so Chandra scored 22 points. Now, I have all the players' scores except for one more. Wait, hold on. The players mentioned are Chandra, Akiko, Michiko, and Bailey. So, I have all four players: Chandra (22), Akiko (11), Michiko (7), and Bailey (14).
Wait a second, I just realized that in the problem, it says Chandra scored twice as many points as Akiko, which we already accounted for. So, now, I have all four players' scores.
Let me just recap to make sure I didn't miss anyone:
- Chandra: 22 points - Akiko: 11 points - Michiko: 7 points - Bailey: 14 points
Now, to find the total points scored by the team, I need to add up all these individual points.
So, let's do that:
Total = Chandra + Akiko + Michiko + Bailey Total = 22 + 11 + 7 + 14
Let me compute that step by step:
First, 22 + 11 is 33.
Then, 33 + 7 is 40.
Finally, 40 + 14 is 54.
So, the total points scored by the team in the final game is 54.
Wait, just to make sure I didn't make a mistake in my calculations, let me add them again:
22 (Chandra) + 11 (Akiko) = 33
33 + 7 (Michiko) = 40
40 + 14 (Bailey) = 54
Yes, that's correct.
Let me also verify each player's score to ensure there were no errors in computation:
- Michiko: Half of Bailey's 14 is 7. Correct. - Akiko: Michiko's 7 + 4 is 11. Correct. - Chandra: Akiko's 11 * 2 is 22. Correct. - Bailey: Given as 14. Correct.
All individual scores check out, so adding them up gives the total.
Therefore, the team scored a total of 54 points in the final game.
I think that's it. I don't see any mistakes in my reasoning or calculations.

PoE Teacher Trace

To solve this problem, I'll break it down step by step and use the given information to find the points scored by each player.
1. Bailey scored 14 points. 2. Michiko scored half as many points as Bailey, so Michiko scored 14 / 2 = 7 points. 3. Akiko scored 4 more points than Michiko, so Akiko scored 7 + 4 = 11 points. 4. Chandra scored twice as many points as Akiko, so Chandra scored 11 * 2 = 22 points.
Now, I'll sum up the points scored by all four players to find the total points scored by the team in the final game of the season.
Total points = Chandra's points + Akiko's points + Michiko's points + Bailey's points Total points = 22 + 11 + 7 + 14 Total points = 54
Therefore, the team scored a total of 54 points in the final game of the season.
$\boxed{54}$

ADS Teacher Trace

Alright, let's tackle this basketball scoring problem step by step. I'm going to break it down and make sure I understand each part before moving on. Hopefully, I won't miss anything!

First, let's read the problem again and identify the key information:

- There are four players: Chandra, Akiko, Michiko, and Bailey. - Each player scored points in the final game. - The relationships between their scores are given in terms of multiples and additions/subtractions. - We're given specific values for some players and need to find the total points scored by the team.

The goal is to find the total points scored by the team in the final game. To do this, we'll need to find each player's score and then add them all together.

Let's list out the information we have:

1. Chandra scored twice as many points as Akiko. 2. Akiko scored 4 more points than Michiko. 3. Michiko scored half as many points as Bailey. 4. Bailey scored 14 points.

Now, let's assign variables to each player's score to make it easier to work with:

- Let C = Chandra's points - Let A = Akiko's points - Let M = Michiko's points - Let B = Bailey's points

We're given that B = 14 points. So we can plug that into the equations we have.

Now, let's write down the relationships as equations:

1. C = 2 * A (Chandra scored twice as many as Akiko) 2. A = M + 4 (Akiko scored 4 more than Michiko) 3. M = (1/2) * B (Michiko scored half as many as Bailey) 4. B = 14 (Given)

So, we can start solving these equations step by step, beginning with the one we know the value for (B) and working our way up to find each player's score.

Let's start with equation 3:

3. M = (1/2) * B We know B = 14, so let's plug that in: M = (1/2) * 14 M = 7

So Michiko scored 7 points.

Now, let's move to equation 2:

2. A = M + 4 We just found that M = 7, so: A = 7 + 4 A = 11

So Akiko scored 11 points.

Next, we'll use equation 1:

1. C = 2 * A We know A = 11, so: C = 2 * 11 C = 22

Therefore, Chandra scored 22 points.

Now we have all the individual scores:

- Chandra (C) = 22 points - Akiko (A) = 11 points - Michiko (M) = 7 points - Bailey (B) = 14 points

To find the total team score, we just need to add up all these individual scores:

Total = C + A + M + B Total = 22 + 11 + 7 + 14

Let's add them step by step:

22 + 11 = 33 33 + 7 = 40 40 + 14 = 54

So the total points scored by the team in the final game is 54 points.

Question

The function $f(x)$ satisfies

$$f(xy) = \frac{f(x)}{y}$$

for all positive real numbers x and y. If $f(30) = 20$, find $f(40)$.

Unperturbed Teacher Trace

Okay, so I have this problem here about a function f(x) that satisfies a certain equation for all positive real numbers x and y. The equation is f(xy) = f(x)/y. And I'm given that f(30) = 20, and I need to find f(40). Hmm, interesting.
Let me try to break this down. First, the function f takes a positive real number and returns another positive real number, right? So, for any x and y, if I multiply them and apply f, it's the same as taking f(x) and dividing it by y.
Wait, so f(xy) = f(x)/y. That seems like a functional equation. I remember that functional equations often have solutions that are power functions, like f(x) = kx^{n} or something like that. Maybe I can assume a form for f(x) and see if it fits the equation.
Let me suppose that f(x) is of the form $f(x) = kx^n$, where k and n are constants. Let's see if that works.
Substituting into the equation: $f(xy) = k(xy)^n = kx^n y^n$. On the other hand, $f(x)/y = (kx^n)/y = kx^n/y$. So, equating both sides: $kx^n y^n = kx^n/y$.
Hmm, that would mean $y^n = 1/y$, which implies $y^{(n+1)} = 1$. But this has to hold for all positive real numbers y, which would only be possible if $n + 1 = 0$, so $n = -1$. So, n would have to be -1.
So, if $n = -1$, then $f(x) = kx^{-1} = k/x$. Let's check if that works.
So, $f(xy) = k/(xy)$. On the other hand, $f(x)/y = (k/x)/y = k/(xy)$. So, yes, they are equal. So, $f(x) = k/x$ is a solution. So, f(x) is a constant divided by x.
Therefore, $f(x) = k/x$. So, that would mean that f(x) is inversely proportional to x.
So, given that $f(30) = 20$, I can find k, and then use that to find $f(40)$.
So, $f(30) = k/30 = 20$. Therefore, $k = 20 \times 30 = 600$. So, $k = 600$.
Therefore, $f(x) = 600/x$.
Therefore, $f(40) = 600/40$. Let's compute that: 600 divided by 40. Well, 40×15 is 600, so $600/40$ is 15. Therefore, $f(40) = 15$.
Wait, that seems straightforward, but let me double-check to make sure I didn't make any mistakes.
So, I assumed that f(x) is of the form k/x, which seems to satisfy the functional equation. Plugging in $f(xy) = k/(xy)$, and $f(x)/y = (k/x)/y = k/(xy)$, which is equal. So, that works.
Then, using $f(30) = 20$, I found $k = 600$. So, $f(x) = 600/x$. Then, $f(40) = 600/40 = 15$. Yep, that seems correct.
Is there another way to approach this problem without assuming the form of f(x)? Maybe by manipulating the functional equation directly?
Let's see. The functional equation is $f(xy) = f(x)/y$ for all positive x and y.
I can try to express f in terms of f at another point. Maybe, set $y = 1$. Then, $f(x \cdot 1) = f(x)/1$, so $f(x) = f(x)$. Hmm, that doesn't help.
Alternatively, set $x = 1$. Then, $f(y) = f(1)/y$. So, $f(y) = f(1)/y$. So, that gives me $f(y) = C/y$, where C is a constant, $f(1)$. So, that shows that f(x) must be of the form C/x. So, that gives me the same result as before.
Therefore, f(x) is C/x, so using $f(30) = 20$, find $C = 600$, so $f(40) = 15$. So, that's consistent.

PoE Teacher Trace

Let's start by analyzing the given functional equation $f(xy) = \frac{f(x)}{y}$ for all positive real numbers x and y.
We are given that $f(30) = 20$ and need to find $f(40)$.
First, let's substitute $x = 30$ and $y = \frac{40}{30} = \frac{4}{3}$ into the functional equation:

$$f\left(30 \cdot \frac{4}{3}\right) = \frac{f(30)}{\frac{4}{3}}$$

Simplifying the left side:

$$f\left(40\right) = \frac{f(30)}{\frac{4}{3}}$$

We know that $f(30) = 20$, so we substitute this value in:

$$f(40) = \frac{20}{\frac{4}{3}} = 20 \cdot \frac{3}{4} = 15$$

Thus, the value of $f(40)$ is $\boxed{15}$.

ADS Teacher Trace

Okay, I've got this problem here about a function f(x) that satisfies the equation $f(xy) = f(x)/y$ for all positive real numbers x and y. I know that $f(30) = 20$ and I need to find $f(40)$. Hmm, I need to figure out how this function behaves based on the given condition and the functional equation.
Let's start by trying to understand the functional equation $f(xy) = f(x)/y$. I notice that the function takes the product of x and y and divides the function value by y. This suggests that the function might have something to do with the reciprocal or the inverse of x or y. I wonder if I can find a general form for f(x) based on this equation.
Maybe I can assume a form for f(x) and see if it fits the functional equation. I'll try a simple form first and see if I can adjust it later. Let's suppose $f(x) = k/x$, where k is a constant. I'll plug this into the functional equation and see if it works.
If I plug $f(x) = k/x$ into the equation $f(xy) = f(x)/y$, I get:
$f(xy) = k/(xy)$
$f(x)/y = (k/x)/y = k/(xy)$
So both sides are equal, which means my assumption $f(x) = k/x$ satisfies the functional equation. That's great! I now know the general form of the function $f(x) = k/x$. I just need to find the constant k using the given condition $f(30) = 20$.
Let's plug in $x = 30$ and $f(30) = 20$ into the equation $f(x) = k/x$:
$20 = k/30$
To find k, I can multiply both sides by 30:
$k = 20 \times 30 = 600$
So the function f(x) is $f(x) = 600/x$. Now I can use this to find $f(40)$:
$f(40) = 600/40 = 15$
Wait a minute, I got $f(40) = 15$. I should double-check my work to make sure I didn't make any mistakes. I'll go through the steps again:
1. I assumed $f(x) = k/x$ and checked if it satisfies the functional equation. 2. I found that it does satisfy the equation. 3. I used the given condition $f(30) = 20$ to find the constant $k = 600$. 4. I then calculated $f(40) = 600/40 = 15$.
Everything seems to check out. I don't see any errors in my calculations or reasoning. I think I've arrived at the correct answer: $f(40) = 15$.
I wonder if I can approach the problem differently to confirm my answer. Maybe I can use the functional equation directly without assuming a form for f(x). I'll try that approach now.
Let's write the functional equation again: $f(xy) = f(x)/y$. I need to find $f(40)$ given $f(30) = 20$. I can try to express $f(40)$ in terms of $f(30)$ by choosing appropriate values of x and y such that $xy = 40$ and $x = 30$.
If I let $x = 30$ and $y = 40/30 = 4/3$, then I have:
$f(30 \cdot (4/3)) = f(30)/(4/3)$
Simplifying the left side:
$f(40) = f(30)/(4/3) = 20/(4/3) = 20 \times (3/4) = 15$
Yes! I get the same result using this method. I can confidently say that $f(40) = 15$. I think I've solved the problem correctly using two different approaches and I'm satisfied with the answer I got.

Question

Solve the following system of equations:

$$3x - 5y = -11,$$
$$7x + 2y = -12.$$

Express your answer as an ordered pair (x, y).

Unperturbed Teacher Trace

Alright, so I have this system of equations to solve. Let me write them down again so I don't get confused.
The first equation is: 3x minus 5y equals -11. So, 3x - 5y = -11.
The second equation is: 7x plus 2y equals -12. So, 7x + 2y = -12.
Okay, so I need to solve for x and y. I remember there are a couple of methods to do this: substitution or elimination. Substitution is where you solve one equation for one variable and then substitute that into the other equation. Elimination is where you manipulate the equations to get rid of one variable by adding or subtracting them.
Looking at these equations, elimination might be a good approach because the coefficients are manageable, and it might be easy to eliminate one of the variables.
Let me see. If I try to eliminate y, I need the coefficients of y in both equations to be the same (or negatives). The first equation has -5y, and the second has +2y. So, if I multiply the first equation by 2 and the second equation by 5, then the coefficients of y would be -10 and +10. That way, when I add the equations, the y terms would cancel out.
Wait, let me check: 2*(3x -5y) = 2*(-11) would give me 6x -10y = -22.
And 5*(7x +2y) = 5*(-12) would give me 35x +10y = -60.
So, now I have:
6x -10y = -22
35x +10y = -60
Now, if I add these two equations together, the y terms will eliminate each other.
Adding them:
6x +35x = 41x
-10y +10y = 0
-22 + (-60) = -82
So, 41x = -82
Therefore, x = -82 divided by 41. Let me compute that.
-82 divided by 41 is -2, because 41*2 is 82, so with the negative sign, it's -2.
So, x is -2.
Now that I have x, I can plug this value back into one of the original equations to find y. Let me choose the second equation because it looks simpler.
The second equation is: 7x + 2y = -12.
Plugging x = -2 into this:
7*(-2) + 2y = -12
Compute 7*(-2): that's -14.
So, -14 + 2y = -12
Now, let's solve for y.
First, I can add 14 to both sides to get 2y by itself.
-14 + 14 + 2y = -12 + 14
Simplify: 0 + 2y = 2
So, 2y = 2
Divide both sides by 2:
y = 1
So, y is 1.
Let me just check these values in both equations to make sure I didn't make any mistakes.
First equation: 3x -5y = -11
Plugging x = -2 and y =1:
3*(-2) -5*(1) = -6 -5 = -11. That's correct.
Second equation: 7x +2y = -12
Plugging x = -2 and y =1:
7*(-2) +2*(1) = -14 +2 = -12. That's correct too.
So, both equations are satisfied with x = -2 and y =1. Therefore, the solution is (-2, 1).
I think that's it. I didn't make any calculation errors, and I checked both solutions, so I'm confident this is correct.
Final Answer The solution to the system of equations is $\boxed{(-2, 1)}$.

PoE Teacher Trace

To solve the system of equations, I will use the method of elimination. First, I will multiply the first equation by 2 and the second equation by 5 to make the coefficients of y in both equations opposites. This gives me:

$$6x - 10y = -22,$$
$$35x + 10y = -60.$$

Next, I will add the two equations together to eliminate y. This gives me:

$$41x = -82.$$

Solving for x, I find that x = -2. Now, I will substitute this value of x back into one of the original equations to solve for y. I will use the first equation:

$$3(-2) - 5y = -11.$$

Simplifying, I get:

$$-6 - 5y = -11.$$

Adding 6 to both sides, I find that -5y = -5. Dividing both sides by -5, I find that y = 1. Therefore, the solution to the system of equations is the ordered pair (-2, 1). $\boxed{(-2, 1)}$

ADS Teacher Trace

Okay, I have to solve this system of equations. Let me write them down again:
3x - 5y = -11
7x + 2y = -12
Hmm, I can use either substitution or elimination. I think elimination might be easier here because I can try to eliminate one of the variables by making the coefficients equal. Let me see... I have two equations with x and y. I need to find a way to eliminate one of them. I'll try to eliminate y first because I have negative coefficients there.
In the first equation, I have -5y and I need to make the coefficient of y the same as the second equation. I can multiply the entire second equation by something to make the coefficients of y the same. I'll see... I have -5y and I want to get the same coefficient as I have in the second equation which I have +2y. I can find the least common multiple of 5 and 2 which is 10. I can make the coefficients of y equal by multiplying the first equation by 2 and the second equation by 5. That way I will have:
First equation multiplied by 2:
2*(3x -5y) = 2*(-11) => 6x -10y = -22
Second equation multiplied by 5:
5*(7x +2y) =5*(-12) =>35x +10y = -60
Now I can add these two equations together to eliminate y:
6x -10y +35x +10y = -22 + (-60)
Simplify:
41x = -82
Oh, I can solve for x now. I just divide both sides by 41:
x = -82 /41
Simplify the fraction:
x = -2
Wait a second, I thought I was going to solve for x and I got x = -2. I should check if I made a mistake somewhere because I have negative coefficients and I ended up with x being a whole number. I'll go back and check my steps. I multiplied the first equation by 2 and the second equation by 5 correctly:
First equation: 6x -10y = -22
Second equation:35x +10y = -60
Adding them together:
6x +35x -10y +10y = -22 -60
41x = -82
x = -2
Okay, I didn't make a mistake there. I'll plug x = -2 back into one of the original equations to find y. I'll choose the second equation because I think the numbers will be smaller:
7x +2y = -12
Substitute x = -2:
7*(-2) +2y = -12
-14 +2y = -12
Now I can solve for y:
2y = -12 +14
2y = 2
y =1
Wait, I got y =1? I need to check if I made a mistake here as I thought I should have a negative coefficient. I'll plug x = -2 and y =1 back into the first equation to verify:
3x -5y = -11
3*(-2) -5*(1) = -11
-6 -5 = -11
-11 = -11
Okay, it checks out. I think I did it right. I just want to make sure I didn't make any calculation errors. I'll go through the steps again:
First equation multiplied by 2:
6x -10y = -22
Second equation multiplied by5:
35x +10y = -60
Adding them:
41x = -82
x = -2
Plugging into the second equation:
7*(-2) +2y = -12
-14 +2y = -12
2y = 2
y=1
Yes, I think I did it correctly. I was just confused because I thought I might have made a mistake with the signs, but I checked it and it's correct. I'll write the final answer as the ordered pair (-2,1). I think I should present it in the box as $\boxed{(-2, 1)}$.

Question

In order for Mateen to walk a kilometer(1000m) in his rectangular backyard, he must walk the length 25 times or walk its perimeter 10 times. What is the area of Mateen's backyard in square meters?

Unperturbed Teacher Trace

Okay, let me try to figure out this problem step by step. So, we have Mateen who wants to walk a kilometer in his rectangular backyard. There are two ways he can do this: either by walking the length 25 times or by walking the perimeter 10 times. We need to find the area of his backyard in square meters.
First, let me note down the important information:
- 1 kilometer = 1000 meters. - He can walk 1000 meters by either: - Walking the length 25 times, or - Walking the perimeter 10 times.
So, let me denote:
- Let the length of the backyard be L meters. - Let the width be W meters.
Therefore, the area we need to find is L * W.
Now, let's analyze the two scenarios given.
First scenario: Walking the length 25 times equals 1000 meters.
So, 25 * L = 1000.
From this, we can solve for L.
L = 1000 / 25.
Let me compute that.
1000 divided by 25 is 40. So, L = 40 meters.
Okay, so the length is 40 meters.
Second scenario: Walking the perimeter 10 times equals 1000 meters.
So, 10 * (Perimeter) = 1000.
First, let's compute the perimeter.
Perimeter of a rectangle is 2*(L + W).
So, Perimeter = 2*(40 + W).
But according to the second scenario:
10 * Perimeter = 1000.
So, 10 * (2*(40 + W)) = 1000.
Let me write that equation.
10 * 2 * (40 + W) = 1000.
Simplify:
20 * (40 + W) = 1000.
Now, let me solve for (40 + W):
40 + W = 1000 / 20.
1000 divided by 20 is 50. So,
40 + W = 50.
Therefore, W = 50 - 40.
Which is W = 10 meters.
So, the width is 10 meters.
Now, we have both length and width.
Length L = 40 meters.
Width W = 10 meters.
Therefore, the area is L * W = 40 * 10.
Compute that:
40 * 10 = 400.
So, the area is 400 square meters.
Let me just double-check my calculations to make sure I didn't make any mistakes.
First scenario:
25 * L = 1000.
L = 1000 / 25 = 40. That seems correct.
Second scenario:
Perimeter = 2*(L + W) = 2*(40 + W).
10 * Perimeter = 1000.
So, 10 * 2*(40 + W) = 1000.
20*(40 + W) = 1000.
40 + W = 50.
W = 10. That also seems correct.
Therefore, area is 40 * 10 = 400. Yep, that's right.
I think that's all. I don't see any mistakes here.

PoE Reasoning Trace

To determine the area of Mateen's backyard, we need to find the dimensions of the rectangular backyard. We know that walking the length 25 times equals 1000 meters, so the length of the backyard is 1000 meters divided by 25, which is 40 meters. We also know that walking the perimeter 10 times equals 1000 meters, so the perimeter of the backyard is 1000 meters divided by 10, which is 100 meters. The perimeter of a rectangle is given by the formula P = 2(L + W), where L is the length and W is the width. We can plug in the values we know to find the width: 100 = 2(40 + W). Simplifying, we get 50 = 40 + W, so W = 10 meters. The area of a rectangle is given by the formula A = L x W, so the area of Mateen's backyard is 40 meters x 10 meters, which is 400 square meters. Therefore, the area of Mateen's backyard is $\boxed{400}$ square meters.

ADS Reasoning Trace

Okay, I'm going to try to solve this problem step by step. Let's read it again carefully and make sure I understand what it's asking for.
Mateen has a rectangular backyard. To walk a kilometer (1000 meters), he has two options: either walk the length of the backyard 25 times or walk the perimeter of the backyard 10 times. I need to find the area of Mateen's backyard in square meters.
Alright, I know I need to find the area of a rectangle. I remember the formula for the area of a rectangle is length multiplied by width (A = l * w). I also know I need to find the length and the width of the backyard to calculate the area.
Now I need to figure out the length and the width of the backyard using the information provided. The problem gives me two different ways to walk a kilometer and I need to translate those into equations to find the length and the width.
First, let's consider the option where Mateen walks the length of the backyard 25 times to cover a kilometer. I can write this as an equation:
25 * length = 1000 meters
From this equation, I can solve for the length of the backyard:
length = 1000 meters / 25 length = 40 meters
Great, I now know the length of the backyard is 40 meters. I need to find the width now. I can use the second option given in the problem to find the width. The problem states that walking the perimeter of the backyard 10 times also covers a kilometer. I can write this as another equation:
10 * perimeter = 1000 meters
I know the perimeter of a rectangle is calculated as twice the length plus twice the width (P = 2l + 2w). I can substitute the length I found earlier into this equation and solve for the width.
First, I can rewrite the perimeter equation using the known length:
10 * (2 * length + 2 * width) = 1000 meters
Let's plug in the length I found earlier (40 meters):
10 * (2 * 40 meters + 2 * width) = 1000 meters
Now I can simplify the equation step by step:
10 * (80 meters + 2 * width) = 1000 meters
Divide both sides of the equation by 10 to make it simpler:
80 meters + 2 * width = 100 meters
Now I can solve for the width:
2 * width = 100 meters - 80 meters 2 * width = 20 meters width = 20 meters / 2 width = 10 meters
Alright, I found the width of the backyard to be 10 meters. I can now use the area formula I mentioned earlier to find the area of the backyard:
Area = length * width Area = 40 meters * 10 meters Area = 400 square meters
Wait a minute, I just want to double-check my work to make sure I didn't make any mistakes. I'll go through the steps again quickly:
1. I converted the distance walked by the length into the equation and solved for the length: - 25 * length = 1000 meters - length = 40 meters
2. I used the perimeter equation and the fact that walking the perimeter 10 times equals a kilometer: - 10 * perimeter = 1000 meters - perimeter = 100 meters
3. I calculated the perimeter of the rectangle using the length I found: - Perimeter = 2 * length + 2 * width - 100 meters = 2 * 40 meters + 2 * width - 100 meters = 80 meters + 2 * width - 2 * width = 20 meters - width = 10 meters
4. I multiplied the length and width to find the area: - Area = 40 meters * 10 meters = 400 square meters
Everything seems to check out. I don't think I made any calculation errors. I can confidently say the area of Mateen's backyard is 400 square meters.

C.5. Adaptive student scores

weight=High $(v(z) > 0)$

DIEC I'm trying to figure out how many miles Jack has driven over the past 9 years. He drives 37,000 miles every four months. First, I need to find out how many sets of four months are in 9 years.
I know there are 12 months in a year, so in 9 years, there are 9 times 12, which is 108 months. Now, I need to divide the total number of months by the number of months in each driving period, which is 4 months.
So, I'll divide 108 months by 4 months per period, which gives me 27 periods.
Now, I'll multiply the number of periods by the miles driven each period to find the total miles driven.
That means 27 periods times 37,000 miles per period equals... let me do the multiplication step by step.
27 times 30,000 is 810,000 miles. 27 times 7,000 is 189,000 miles. Adding those together gives me 810,000 + 189,000 = 999,000 miles.
Wait, I think I made a mistake in the multiplication earlier. Let me try it again more carefully.
37,000 miles per four months times 27 periods.
I can break it down as (30,000 + 7,000) times 27. 30,000 times 27 is 810,000. 7,000 times 27 is 189,000. Adding those together gives me 810,000 + 189,000 = 999,000 miles.
Wait, I think I'm still getting the same result, but I'm not sure if it's correct. Maybe I should try another method to verify my answer.
Alternatively, I can calculate how many miles Jack drives in a year and then multiply by 9 years.
He drives 37,000 miles every four months. There are 12 months in a year, so I'll divide 12 by 4 to find out how many driving periods are in a year.
12 divided by 4 is 3 periods.
So, in one year, he drives 37,000 times 3, which is 111,000 miles per year.
Now, multiplying the annual miles by 9 years gives me 111,000 times 9.
111,000 times 9 is... let's see: 100,000 times 9 is 900,000, and 11,000 times 9 is 99,000. Adding those together gives me 999,000 miles.
Okay, now I'm confident that the correct answer is 999,000 miles. </think>
To determine how many miles Jack has driven over the past 9 years, we'll follow these steps:
1. **Calculate the number of four-month periods in 9 years:** - There are 12 months in a year. - In 9 years: $9 \times 12 = 108$ months. - Number of four-month periods: $\frac{108}{4} = 27$ periods.
2. **Determine the total miles driven:** - Jack drives 37,000 miles every four months. - Total miles: $27 \times 37,000 = 999,000$ miles.
Final Answer: $\boxed{999000}$

weight=Low $(v(z) < 0)$

First, I recognize that the spaceship's speed is halved for every additional 100 people beyond the initial 200.
raiding raiding

Figure 7. An adaptive student filters traces based on downstream gradient alignment.