CAN DOMAIN SYNERGY MISLEAD? MEASUREMENT CONFOUNDS IN MULTI-DOMAIN RLVR INTERACTION ESTIMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-domain RLVR post-training mixes mathematical reasoning, coding, agentic tool use, and general knowledge into one training run. Existing accounts of why particular domain combinations help fit an additive-plus-pairwise surrogate to a single outcome; none decomposes interaction by *order* beyond the pairwise term, and none tests whether the recovered structure *transports* from the suite used to select a mixture to the suite the model is deployed against. We run a paired, fixed-regime audit over the full coalition lattice of four RLVR domains, holding a source suite and an independent transfer suite fixed throughout. Its central finding is a measurement confound with an exact component that holds for any budget-matched coalition grid: proportional per-coalition budget-splitting ties each domain's update to coalition size, so even a null in which domains do not interact carries a nonzero Möbius (Harsanyi) dividend at every order — a finite difference of each domain's budget-response curve. Its size is empirical; uncorrected, our lattice (GOPO, Qwen3-1.7B) shows significant order-3 and order-4 interaction on both suites. We subtract the null two ways — from single-domain budget-response curves, and with a three-domain lattice at fixed per-domain budget — and add a seed-and-item bootstrap that changes the variance model rather than the estimand. In this setting, at most one order-≥ 3 dividend survives multiplicity correction on transfer under any of these routes, and source is unresolved; a lattice built by merging single-domain experts, with no joint training, reproduces 6 of the 8 uncorrected source survivors and, as a positive control, loses them under the same correction. A second algorithm and model, which cannot be corrected the same way, does not reproduce the uncorrected order profile. A pairwise selection surrogate then loses essentially nothing, while the transfer-optimal coalition is not identifiable from a single-seed grid. We also measure, on a trained RLVR lattice, the Fisher-proportional-curvature condition the Hessian-alignment domain-generalization literature invokes: across our four domains it is inconsistent with a single shared constant. Finally, GIFT, a Bayesian experimental-design estimator, recovers the lattice from a fraction of the budget; its A-optimal rule beats random at every tested budget, on the uncorrected and on the budget-corrected lattice alike, while the guarantee-carrying D-optimal rule does not.

1 INTRODUCTION

Multi-domain RLVR post-training mixes mathematical reasoning, coding, agentic tool use, and general knowledge into one training run (Hamidieh et al., 2026; Liang et al., 2025; Wang et al., 2026; Ma et al., 2026). The direct accounts we can identify of why particular domain combinations help expose synergy through pairwise transfer measurements (Yang et al., 2026), a quadratic mixture-weight surrogate (Liang et al., 2025) (Hamidieh et al. (2026)'s related but distinct token-co-occurrence account, §2, truncates the same way for a different reason), or a small number of explicit domain combinations evaluated on one outcome (Li et al., 2025). Li et al. (2025)'s three-domain grid does report non-additivity at order 3 (§2), but none of this work decomposes interaction by order in a controlled way, and none tests whether one fixed interaction structure transports from the suite used to select a coalition to the suite the resulting model is deployed against. Broader cross-domain

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

transfer studies (Cheng et al., 2025; Hu et al., 2026) report transfer heterogeneity but do not model interaction order either.

This matters concretely. A coalition chosen because it looks best on the suite used to tune it can be the wrong choice on deployment. A coalition-grid audit built to detect that risk is itself vulnerable to *manufacturing* the high-order structure it goes looking for, through nothing more than how it allocates training budget across coalition sizes. This second point is the paper's central, initially unwelcome finding, and it is measured directly (§5.2), not assumed.

We confront the two closest works directly. Li et al. (2025) run the full coalition grid across three RLVR domains; applying the Möbius operator of §3 to their Table 12 gives a three-way dividend of $+8.51$ points on their ALL column (reproduced in `scripts/verify_theory.py`), so three-domain non-additivity in RLVR is not, on its own, a new observation. Wang et al. (2026) compare pooled joint training against expert-then-merge construction at comparable aggregate performance. What neither does is hold two independent evaluation regimes fixed and ask whether the *structure* of interaction transports between them, and neither budget-matches across coalition sizes, so neither is exposed to the confound identified here.

Contributions.

1. **A paired fixed-regime audit protocol, and a measurement confound in it with an exact component.** Holding a source and a transfer suite fixed across the full coalition lattice (§4), we show that budget-matching by coalition size gives the non-interacting null a nonzero Möbius dividend at every order — exactly a finite difference of each domain's budget-response curve, for any budget-matched grid (§3.2); its size is empirical. We subtract that null two ways — from measured single-domain budget-response curves (correction 1) and with a three-domain lattice at fixed per-domain budget, which retests its single order-3 coalition (condition E) — and add a seed-and-item bootstrap that changes the variance model, not the estimand (correction 3). *In our setting* (GOPO, Qwen3-1.7B, one primary seed, $|D| = 4$), no route leaves **more than one order-≥ 3 dividend surviving multiplicity correction on transfer**, and source is unresolved; a lattice merged from single-domain experts (condition C1) reproduces most uncorrected source survivors without joint training and loses them under correction 1 (§5.2). A second algorithm and model does not reproduce the uncorrected order profile and cannot be corrected the same way (§5.6): the mechanism is general, this null is not yet. At $|D| = 4$ a pairwise selection surrogate then loses essentially nothing, while the transfer-optimal coalition is not identifiable from a single-seed grid (§5.3).
2. **An RLVR-specific curvature account, and the first direct measurement of the generic condition it draws on.** For GOPO we derive in closed form (§3.3, Appendix A.3) that evaluation-risk curvature is a reward-weighted second moment of the score, vanishing when reward does not depend on the sampled completion. This licenses the first direct measurement, on a trained RLVR lattice, of the generic proportional-curvature condition the domain-generalization Hessian-alignment line (Hemati et al., 2023) invokes for cross-regime sign agreement. Across our four domains it is **inconsistent with a single shared constant** (§5.4); our own reward-weighted antecedent — the object the derivation actually concerns — remains unmeasured and is stated as such.
3. **GIFT, a scalable estimator whose target-dividend family grows polynomially, not as $2^{|D|}$, with domain count at a fixed order ceiling.** On the trained lattice its A-optimal acquisition rule recovers the *uncorrected* dividends from a fraction of the training budget, beating random selection at every tested budget on both suites, while the D-optimal objective carrying a proven near-optimality guarantee does not; the same holds on the budget-corrected lattice (§5.5). We claim no novelty in the design machinery, an instantiation of published goal-oriented optimal design (Attia et al., 2018; Alexanderian, 2021); the contribution is the target functional and which of two standard objectives works here.

2 Related Work

Multi-domain RLVR mixture and synergy. MoDoMoDo (Liang et al., 2025) models domain interaction as a quadratic surrogate over mixture proportions — an explicit, stated truncation it does

not test for sufficiency. Frontier reports take the same shape at production scale: Kimi K2.5 (Kimi Team, 2026) trains its abilities (knowledge, reasoning, coding, agentic) jointly and reports the resulting synergy as a single aggregate gain, with no interaction-order decomposition, fixed transfer suite, or mixture ablation. Domain-Aware Scaling Laws (Hamidieh et al., 2026) is a related but distinct object: its pairwise term is a softmin-gated bonus fit on logged token co-occurrence counts across pretrained models, not a regression over mixture proportions within one model family. TAC (Yang et al., 2026) builds a curriculum from pairwise transferability signals. Breaking Barriers (Hu et al., 2026) and GURU (Cheng et al., 2025) report single-domain and cross-domain transfer heterogeneity across many domains without decomposing it by interaction order. Li et al. (2025) and Wang et al. (2026) are the two closest works, confronted in §1. The Lie-bracket account of non-commutativity between domain gradient flows (Rukhovich et al., 2024) and continuous mixture-weight optimization over a simplex (Liang et al., 2025) are different mathematical objects from the binary present/absent coalition game studied here.

Any-order interaction estimation. Shapley-Taylor (Sundararajan et al., 2020), SHAP-IQ (Fumagalli et al., 2023), SVARM-IQ (Kolpaczki et al., 2024), KernelSHAP-IQ (Fumagalli et al., 2024), and PolySHAP (Fumagalli et al., 2026) establish mature machinery for recovering arbitrary-order interaction structure from a value function; ShaplEIG (Rundel et al., 2026) adds Bayesian design for the costly-retrain regime. GIFT's contribution relative to this line (§4.1) is a multi-target acquisition rule for a *family* of Möbius dividends simultaneously, rather than ShaplEIG's single-target Shapley value — an instantiation of goal-oriented Bayesian optimal experimental design (Attia et al., 2018; Alexanderian, 2021), not new design machinery.

Coalition/set-function templates and task grouping. Data Shapley (Ghorbani & Zou, 2019), randomized subset retraining (Lin et al., 2022), and Datamodels (Ilyas et al., 2022) treat training units as coalition-game players; task grouping (Standley et al., 2020; Fifty et al., 2021) and higher-order task affinity (Li et al., 2023) already show pairwise relations miss group effects in ordinary multi-task learning. None holds two independent evaluation distributions fixed and tests transport of the recovered structure between them; the closest, ProxySPEX (Butler et al., 2025) and a task-attribution kernel surrogate (Zhang et al., 2026), stop at second order, model one target task, and validate on continuous-control RL.

Hessian-alignment domain generalization. Hemati et al. (2023) prove source/target Hessian mismatch bounds a formal transfer measure and motivates a regularizer family (Ramé et al., 2022; Shi et al., 2022); Chen et al. (2025) extend this to a multi-source bound. That literature motivates but never directly measures the proportional-curvature condition Proposition 1 states; §5.4 reports the first direct measurement of it on a trained RLVR lattice.

3 THE AUDIT AND ITS CONFOUND

Let D be a finite set of RLVR training domains. For $S \subseteq D$, let $\theta_S = \theta_0 + \Delta\theta_S$ be the parameters after RLVR on the union of domains in S under a fixed total compute budget B split proportionally across active domains, so subsets of different sizes stay comparable. We fix **two** evaluation distributions independent of S — a source suite P_{src} (used to select a coalition) and a transfer suite P_{tr} (standing in for deployment) — rather than an "in-domain" suite, whose meaning would change with S. The value function and interaction operator are

$$v_P(S) = -\big(R_P(\theta_S) - R_P(\theta_0)\big), \quad R_P(\theta) = \mathbb{E}_{x\sim P}[\ell(\theta; x)]; \quad \Delta[v](S) = \sum_{T\subseteq S} (-1)^{|S|-|T|} v(T), \tag{1}$$

where R_P is the task risk (so v_P is an improvement, equivalently a score gain $\mathrm{score}_P(\theta_S) - \mathrm{score}_P(\theta_0)$ as tabulated in Appendix A.9), and Δ is the Möbius transform, coinciding with the Harsanyi dividend of cooperative game theory (Hammer & Rudeanu, 1968; Owen, 1972; Rota, 1964). A nonzero $\Delta[v_P](S)$ at $|S| \geq 3$ is order-≥ 3 interaction. Each suite has one score per training domain — four in all, math pooling its three transfer benchmarks (Table 1) — so v_P and its dividends are computed for every coalition on each of the four domain scores. The Benjamini–Hochberg family is the 60 dividends of one suite (4 scores $\times$ 15 coalitions), run separately per suite at $q = 0.1$; the order-≥ 3 subset that Table 2 reports is $4 \times 5 = 20$ of those 60.

3.1 A CONDITIONAL NULL, USED ONLY AS A SCAFFOLD

Lemma 1. *Under a second-order Taylor expansion of R_P around θ_0, if joint training on S produces $\Delta\theta_S = \sum_{d\in S} \delta_d$ (the sum of per-domain updates at the budget the coalition allocates each domain), then $\Delta[q_P](S) = 0$ for every S with $|S| \geq 3$ and every P, where q_P is the quadratic model.*

The proof (Appendix A.2) is the Hammer–Rudeanu/Owen correspondence: an additive $\Delta\theta_S$ makes $q_P(S)$ a degree-≤ 2 pseudo-Boolean function, whose Möbius transform vanishes above order 2. We use Lemma 1 **only as a diagnostic scaffold** for interpreting a measured deviation, since both of its licensing conditions fail on our runs: the FLOP-ratio condition for the quadratic expansion (Meterez et al., 2026) misses by roughly four orders of magnitude, and the weight-space linearity gate fails on every domain (§5.6, Appendix A.6); Figure 2 contrasts the two worlds. The corrected result of §5.2 is expansion-free.

3.2 THE BUDGET-ALLOCATION CONFOUND

Splitting B proportionally to $|S|$ gives each domain the per-domain budget $b(S) = B/|S|$, so δ_d is *not* independent of $|S|$ as Lemma 1 assumes. Write $f(S) = 1/|S|$ for $S \neq \emptyset$ and $f(\emptyset) := 0$; because f depends on coalition size alone, its Möbius transform $\Delta[f](S) = \sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j} f(j)$ for $k = |S|$ is the same for every coalition of a given size, and equals

$$\Delta[f](S) \;=\; (-1)^{|S|-1} H_{|S|} \;=\; -1.500,\; +1.833,\; -2.083,\; +2.283,\; -2.450 \quad \text{at } |S| = 2, \ldots, 6, \tag{2}$$

(H_k the k-th harmonic number), independent of $|D|$: alternating in sign, and rising slowly with order ($1.5 \to 2.08$ from order 2 to 4).

What this costs the audit is not equation 2 itself but the dividend of the *null in which domains do not interact at all*. Let $\varphi_d(b)$ be domain d's measured single-domain response at per-domain budget b, and $\tilde{v}_P(S) = \sum_{d\in S} \varphi_d(B/|S|)$ that null under budget-matching. Writing $g_d(j) = \varphi_d(B/j)$, linearity of Δ gives, for $|T| = k$,

$$\Delta[\tilde{v}_P](T) \;=\; \sum_{d\in T} \Delta^{k-1} g_d(1), \tag{3}$$

the $(k-1)$-st forward difference of each member domain's own budget-response curve. It vanishes at order $k \geq 3$ only if every g_d is affine in j; otherwise **a budget-matched grid reports nonzero dividends at every order with no interaction present** — the exact part, and what correction 1 subtracts. Additivity is the definition of this null, not an assumption about training: "no interaction" means each member contributes what it contributes when trained alone at the budget it received, which is $\tilde{v}_P$; whatever joint training adds, from shared optimizer state or from capability transfer, is what $v_P - \tilde{v}_P$ measures. The identity itself is an elementary finite difference (Appendix A.2); what matters is that budget-matched RLVR grids fall exactly under it. It does *not* imply growth with order: for a response linear in b, equation 3 collapses to $(-1)^{k-1} B\,\bar{\alpha}_T$ ($\bar{\alpha}_T$ the mean slope over T), alternating but flat in k. Magnitude and order profile belong to the measured φ_d, so we take them from the budget sweep, not from equation 2. A leading-order optimizer argument (Appendix A.6) models why single-domain training responds to b at all; it is not load-bearing, and its displacement-level form is shown there to be unlicensed at the displacements these runs reach. Fixing the *per-domain* budget would make g_d constant in j, so $\Delta[\tilde{v}_P] = 0$, but answers a different question (interaction structure in isolation) than the one a practitioner faces (one coalition under one total compute allocation), which is the estimand this paper audits. Under that estimand the dilution term is part of the value a practitioner actually obtains; what it cannot support is reading a nonzero high-order dividend as capability-level interaction, which is the reading the literature in §2 invites. We therefore budget-match and subtract equation 3 from data, two ways, with further routes that vary the variance model or remove joint training instead (§5.2).

3.3 PAIRWISE INTERACTION IS CURVATURE-DEPENDENT

Under additivity, $\Delta[q_P](\{d, d'\}) = -\delta_d^\top H_P \delta_{d'}$, so by linearity of Δ,

$$\Delta[v_{\mathrm{src}}](\{d, d'\}) - \Delta[v_{\mathrm{tr}}](\{d, d'\}) = -\,\delta_d^\top (H_{\mathrm{src}} - H_{\mathrm{tr}})\, \delta_{d'}. \tag{4}$$

Proposition 1. *If $H_{\mathrm{src}} = c \cdot H_{\mathrm{tr}}$ on $\mathrm{span}\{\delta_d\}$ for a scalar $c > 0$, pairwise interaction agrees in sign between source and transfer for every domain pair.*

This is sufficient for sign agreement, not necessary. Vanilla GRPO's clipped surrogate admits no closed-form Hessian, so we work through Group Orthogonalized Policy Optimization (GOPO) (Wang, 2026), whose smooth quadratic penalty has constant curvature μ in the policy ratio; Appendix A.3 derives

$$\begin{aligned} H_P^{\mathrm{gopo}}(\theta_0) &= \mu\, F_P(\theta_0) + H_P^{\mathrm{true}}(\theta_0), \\ H_P^{\mathrm{true}}(\theta_0) &= -\mathbb{E}_{x\sim P,\, y\sim\pi_{\theta_0}}\big[r(y)\,\big(S(y)S(y)^\top + \nabla_\theta^2 \log \pi_{\theta_0}(y)\big)\big], \end{aligned} \tag{5}$$

with $S(y) = \nabla_\theta \log \pi_{\theta_0}(y)$ the score and F_P the Fisher information. μF_P is curvature the trust-region penalty itself contributes and must not appear in equation 4; the remainder H_P^{true} is a *reward-weighted second moment of the score*, vanishing when r does not depend on the sampled completion — distinct in kind from supervised domain generalization, where curvature depends on a static distribution rather than on which trajectories a policy induces. Proposition 1's literal antecedent is a statement about H_P^{true}; §5.4 measures the weaker *generic* Fisher-proportionality condition that the Hessian-alignment literature (Hemati et al., 2023) operationalizes, and refutes it.

4 Method

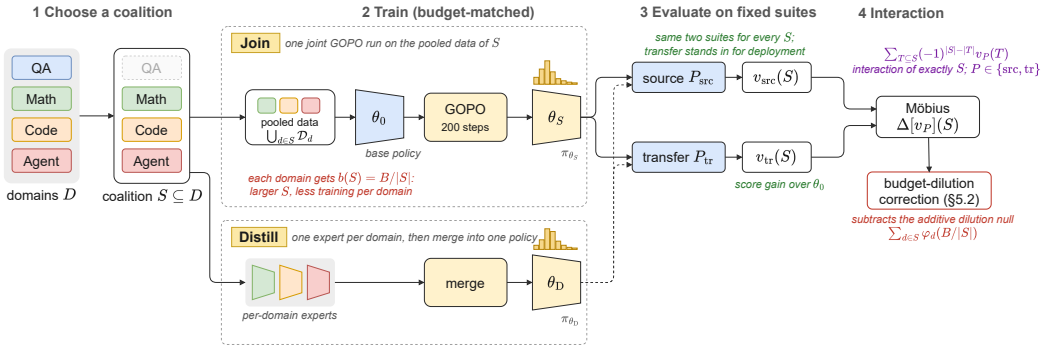


Figure 1: Coalition audit pipeline. A coalition $S \subseteq D$ is trained under a fixed total budget, so each member domain receives $b(S) = B/|S|$ (red); the resulting policy is scored on the same two frozen suites for every S, and the two score gains feed the Möbius operator $\Delta[v_P](S)$, from which the budget-dilution null equation 3 is subtracted (§5.2). Yellow marks objects produced by training, blue marks objects held fixed across the lattice (θ_0 and the two suites), white marks quantities derived from them. Solid arrows are the **Join** arm (pooled joint RLVR), dashed arrows the **Distill** arm, here a task-vector merge of single-domain experts each trained at $B/|S|$ (condition C1, §5.2).

Protocol. Figure 1 shows the pipeline. Given D, fixed $P_{\mathrm{src}}, P_{\mathrm{tr}}$, and budget B: if $|D| \le 5$, enumerate every non-empty $S \subseteq D$; else select a query budget of B_q coalitions adaptively with GIFT (§4.1). For each S, construct $\Delta\theta_S$ by **Join** (pooled joint RLVR on the union of domain data — the only construction studied in the pairwise-synergy literature (Liang et al., 2025; Hamidieh et al., 2026)) or by **Distill** (merging single-domain experts; condition C1, §5.2), measure $v_{\mathrm{src}}(S), v_{\mathrm{tr}}(S)$ with paired eval-item bootstrap intervals, compute $\Delta[v_P](S)$, correct the budget-dilution confound, and apply Lemma 1's null test and the decision-relevance criterion below, both fixed before any run (Appendix A.4).

Decision-relevance and transport. Let $\hat{v}_2^{(P)}$ be a degree-≤ 2 pseudo-Boolean surrogate fit from singleton and pair data for regime P — the binary-corner restriction of Liang et al. (2025)'s quadratic mixture-weight model — and $\mathrm{top1}_P(\cdot) = \arg\max_S$ over a candidate set.

Definition 1 (Transport failure, decision change, selection regret)**.** Transport failure: $\mathrm{top1}_{\mathrm{src}}(v_{\mathrm{src}}) \neq \mathrm{top1}_{\mathrm{tr}}(v_{\mathrm{tr}})$. Decision change: $\mathrm{top1}_{\mathrm{tr}}(\hat{v}_2^{(\mathrm{tr})}) \neq \mathrm{top1}_{\mathrm{tr}}(v_{\mathrm{tr}})$. *For a rule producing $\hat{S}$,* regret *on transfer is* $\mathrm{Reg}(\hat{S}) = v_{\mathrm{tr}}(S^\star_{\mathrm{tr}}) - v_{\mathrm{tr}}(\hat{S})$, $S^\star_{\mathrm{tr}} = \mathrm{top1}_{\mathrm{tr}}(v_{\mathrm{tr}})$. *We report* Reg *for three rules: order-truncation alone ($\hat{S} = \mathrm{top1}_{\mathrm{tr}}(\hat{v}_2^{(\mathrm{tr})})$), suite-transport alone ($\hat{S} = \mathrm{top1}_{\mathrm{src}}(v_{\mathrm{src}})$),*

*and the **compound** rule ($\hat{S} = \mathrm{top1}_{\mathrm{src}}(\hat{v}_2^{(\mathrm{src})})$) — a pairwise surrogate fit on source and deployed on transfer, the scenario §1 opens with.*

With $|D| = 4$, a degree-≤ 2 surrogate has $1+|D|+\binom{|D|}{2} = 11$ coefficients fit on the 11 coalitions of size ≤ 2, so it interpolates them exactly and $v^{\text{full}}(S) - \hat{v}_2(S) = \sum_{|T|\geq 3,\, T\subseteq S} \Delta[v](T)$: **a decision change cannot occur unless some order-≥ 3 dividend is nonzero.** Transport failure carries no such dependency, so we treat it as the primary endpoint.

4.1 GIFT: A SCALABLE ESTIMATOR

Exhaustive enumeration costs $2^{|D|} - 1$ training runs. GIFT places a GP prior over $v : 2^D \to \mathbb{R}$ and selects the next coalition to train by expected information gain about the target family $\{\Delta[v](T) : |T| \leq k+1\}$, extending ShaplEIG (Rundel et al., 2026) from the Shapley value to the Möbius functional; a fixed-order dividend has support on only $\leq 2^{|T|}$ subsets, so its GP posterior is closed-form (Appendix A.5). The natural D-optimal objective $\mathrm{EIG}_{\text{logdet}}(S^*) = \frac{1}{2}\log\det\Sigma_\Delta^{\text{before}} - \frac{1}{2}\log\det\Sigma_\Delta^{\text{after}(S^*)}$ carries a proven $(1-1/e)$ near-optimality guarantee (Krause & Guestrin, 2005) for greedy mutual-information reduction against its own objective — not a guarantee that log-det reduction proxies dividend-recovery error. Against synthetic ground truth ($|D|$=6) it *loses to random selection* (0.202 vs. 0.176 mean dividend error), spending its entire budget on $|S|$=3 coalitions. The A-optimal alternative $\mathrm{EIG}_{\text{trace}}(S^*) = \frac{1}{2}\mathrm{tr}\,\Sigma_\Delta^{\text{before}} - \frac{1}{2}\mathrm{tr}\,\Sigma_\Delta^{\text{after}(S^*)}$ beats random on 8 of 8 seeds (0.109 vs. 0.176) and is our release's default, but carries *no* equivalent guarantee, which we state as an open gap rather than a resolved one (Appendix A.5). We fix $k+1 = 3$, matching the highest order at which Li et al. (2025)'s grid reports a nonzero dividend. GIFT is an enabling tool, not load-bearing for Contributions 1–2.

5 EXPERIMENTS

5.1 SETUP

Table 1: Condition-A domains, training sets, and evaluation suites as actually used. Three of four rows deviate from an earlier planned design; every deviation is dated in `prereg.json` and detailed in Appendix A.10.

Domain	Train	Source suite	Transfer suite
General QA	MMLU-Pro (2k)	MMLU-Pro	SciKnowEval[a]
Mathematics	DAPO-Math (2k)	DAPO-Math	AIME24, AIME25, OlympiadBench
Coding	MBPP (474)[b]	MBPP+	HumanEval+[b]
Multi-turn agentic	BFCLv3 live/multiple/parallel (1652)[c]	BFCLv3 simple (400)	BFCLv3 live-simple (258)[c]

[a]GPQA-Diamond is hub-gated: a documented reduction of $P_{\mathrm{tr}}(\mathrm{qa})$, not a silent substitution. [b]CodeForces (θ_0 0.010 pass rate, abandoned before any run), CodeElo and LCBv6 were dropped; MBPP→HumanEval is a milder shift. [c]BFCLv3 MT base ($\approx$ 192 epochs at budget) was replaced with AST-verifiable single-turn categories.

Our primary lattice, **condition A**, trains all 15 non-empty subsets under Join with GOPO on Qwen3-1.7B (its own validated setting (Wang, 2026)): 200 steps $\times$ 4 rollouts, 19,200 prompt-draws per coalition, seed 0, with θ_0 scoring $P_{\mathrm{src}} = 0.5230$, $P_{\mathrm{tr}} = 0.5709$. Two further lattices each vary one factor: **condition E** (§5.2) fixes the per-domain budget so coalition size never divides it, and **condition B** (§5.6) swaps GOPO for clipped GRPO on Qwen2.5-1.5B. Three training seeds $\{0, 1, 2\}$ were ultimately completed for every condition-A coalition; the nested bootstrap (§5.2, correction 3) uses all three — the one departure from the one-seed compute policy, adopted because it moves every count in the conservative direction (more uncertainty, fewer discoveries) and seed-1 provenance was audited (15/15 pass, Appendix A.7). Five measurement-path defects were found and corrected first — the largest, a single-draw evaluation issuing k identical decodes, inflated one domain's measured transport failure by $\approx 2.2\times$; θ_0 reproduced to ± 0.001 on three of four domains after every fix (`code`'s changed under a corrected extractor; diagnostics: Figure 4).

5.2 THE ORDER-≥ 3 RESULT: UNCORRECTED, THEN CORRECTED

Uncorrected, the lattice looks like higher-order synergy. Every Möbius dividend (the 60-dividend per-suite family of §3) carries a paired-item bootstrap CI (1,000 resamples) and Benjamini–Hochberg FDR at the pre-registered $q = 0.1$. Mean $|\Delta|$ by order rises monotonically (source $0.0097/0.0140/0.0242/0.0542$, transfer $0.0075/0.0135/0.0240/0.0476$ for orders 1–4) and **8/20 order-≥ 3 dividends survive FDR on source, 6/20 on transfer** (Table 2, first row). Read naively, this is what the multi-domain RLVR literature would want: significant interaction at orders 3 and 4, on both suites, growing with order.

On transfer, it does not survive correction. Two routes subtract the non-interacting null equation 3; a third widens the variance model; a fourth removes joint training (details: Appendices A.6–A.8). *(1) Nonparametric budget-null.* The pre-registered budget sweep trains each domain alone at $b \in \{0.25, 0.5, 0.75, 1.0\}B$, and $\{1, 0.5, 0.25\}B$ are exactly what $|S| = 1, 2, 4$ realize, so $\tilde{v}_P$ of equation 3 is computable from single-domain runs alone (order-1 residual dividends are identically zero by construction, verified). Dividends of $v_P - \tilde{v}_P$: transfer falls $6/20 \to \mathbf{0/20}$, source $8/20 \to \mathbf{6/20}$, with the same surviving set under linear and log interpolation at $B/3$; with $B/3$ measured directly, $1/20$ and $2/20$. *(2) From-scratch design (condition E).* A second lattice over $\{\text{agent}, \text{code}, \text{qa}\}$ gives every member domain the full per-domain budget $B/2$ regardless of coalition size, so $|S|$ never divides the budget and no $1/|S|$ artifact is injected. Prediction committed before any run: if the order-3 dividend survives here as under condition A, the confound cannot explain condition A. It *vanishes* — transfer $1/3 \to \mathbf{0/3}$ on the shared coalition, with power adequate for condition A's measured effect there. Condition E carries its own one-directional total-compute confound (larger coalitions train longer) and covers three domains, so it rules out the confound *mechanism* on the coalitions it spans rather than serving as co-equal evidence on `math`. *(3) Nested seed-and-item bootstrap.* With three seeds, resampling seeds and items jointly (rather than items alone, which is blind to training-seed variance) gives transfer $6/20 \to \mathbf{1/20}$, source $8/20 \to \mathbf{2/20}$; the three survivors are all `math` dividends (Appendix A.7). This route subtracts nothing: it is run on the *uncorrected* lattice, so the two mechanisms — budget dilution and an item-only variance model — are not separately identified here, and neither count alone should be read as the size of the confound. *(4) No joint training (condition C1, Distill).* Coalitions built as $\theta_0 + \sum_{d \in S} \delta_d (B/|S|)$ from experts trained alone keep the $1/|S|$ dilution but share no optimizer state, batch, or advantage normalization (predictions pre-registered). Uncorrected, 11/20 source dividends survive (6 of A's 8); correction 1 removes all ($\mathbf{0/20}$). Transfer shows none even uncorrected ($0/20$; $1/20$ corrected), and $|\Delta|$ rises with order under a size-dependent merge coefficient $\lambda = 1/|S|$ ($1.00/1.26/1.80/2.73$, transfer) — not unique to λ: C2, whose displacement grows with $|S|$, rises on source too (Appendix A.8); it is flat only where $\|\Delta\theta_S\|$ stays fixed (C1). Two of route 3's three `math` survivors (both source) recur with the same sign.

Table 2: Order-≥ 3 Möbius dividends surviving BH-FDR at $q = 0.1$ (20 of the 60 BH-FDR tests per suite: 4 domains $\times$ 5 coalitions of order ≥ 3), and the regret decomposition of Definition 1. Routes 1–2 use seed 0; route 3 adds seeds 1–2 and changes only the variance model. After correction 1, order-3 MDE_{80} is 0.036–0.095 on transfer and 0.048–0.063 on source, 0.9–1.3$\times$ its uncorrected value (Appendix A.6): a zero count rules out effects of at least that size, not all effects.

order-≥ 3 dividends surviving FDR	tr	src	note
Uncorrected (item bootstrap, seed 0)	6/20	8/20	$\|\Delta\|$ grows with order
Correction 1: nonparametric budget-null	**0/20**	6/20	+noise sensitivity: tr 0/20, src 4/20; $B/3$ meas. 1/20,2/20
Correction 2: fixed per-domain budget (cond. E)	**0/3**†	2/3†	†one order-3 coalition; src flips sign
Correction 3: seed$\times$item bootstrap	**1/20**	2/20	3 survivors, all `math`
Cond. C1 (merge): uncorr. $\to$ corr. 1	0/20$\to$1/20	11/20$\to$**0/20**	no joint training; $B/3$ measured
Regret (Def. 1), nested seed$\times$item resampling, 1998 draws:			
order-truncation alone‡	nonzero 0.05%		median 0, mean 3×10^{-10}
suite-transport alone	nonzero 71.5%		median 0.0025, $p_{97.5}$ 0.023
compound (src$\to$tr)	nonzero 68.7%		median 0.0019

†Condition E's one shared coalition; its `src` entries flag *different* benchmarks from A's. $^\ddagger \approx 0$ by construction at $|D| = 4$ (§5.3), not independent evidence.

Why raw magnitude-by-order is not evidence. Run-to-run training nondeterminism at fixed seed, measured for the first time from three accidental replicate pairs (a thin basis, stated as such), is $\sigma \approx 0.004$–0.005. A dividend of order r is an alternating sum over 2^r values, $2^r - 1$ of which carry this noise, so pure noise alone produces a dividend spectrum that *grows* as $\sqrt{2^r - 1}$: the floor rises $3.9\times$ from order 1 to 4. The measured transfer mean $|\Delta|$ clears that floor by 1.9–$3.9\times$ across orders and rises $6.3\times$ overall (Figure 3), so the raw "growth with order" is a mix of real structure and transform-amplified noise that magnitude alone cannot separate — only the FDR step does. The closed-form budget correction q^{budget} cannot be used here either: at the displacement these runs reach ($\|\delta\| = 0.214$) it overshoots the quantity it corrects by $279\times$ (Appendix A.6), which is why corrections 1–3 are expansion-free.

Summary. The routes are not independent: corrections 1 and 3 both build on the seed-0 condition-A lattice, and only condition E retrains. On transfer, **correction 1 leaves 0/20 and condition E leaves 0/3** on its shared coalition; **the seed-inclusive bootstrap leaves 1/20** — a `math` three-way ($\{$agent,math,qa$\}$; $\Delta = -0.0199$, CI_{95} $[-0.033, -0.006]$). As registered, nothing survives FDR under either budget-focused route, but correction 1 does not remove everything: the four former three-way survivors shrink to 40–77% of their size, while the two order-4 ones keep theirs (`tr:code` $+0.163 \to +0.136$, `tr:math` $+0.038 \to +0.041$) and drop out only under the 60-test FDR step; with $B/3$ measured, `tr:code`'s reaches $+0.211$ and survives (Appendix A.6). We do not claim a categorically clean transfer null. The source suite is *unresolved, not settled at zero*: correction 1 leaves 6/20, correction 3 leaves 2/20, and condition E flags different benchmarks and lacks `math`.

5.3 Decision-level consequence and non-identifiability

At $|D| = 4$ a degree-≤ 2 surrogate interpolates the 11 size-≤ 2 coalitions exactly, so order-truncation regret *is* $\sum_{|T|\geq 3} \Delta[v](T)$ by construction, ≈ 0 whenever the order-≥ 3 null holds; resampling it (Table 2, lower block: nonzero in 0.05% of 1998 draws, vs. 71.5% for suite-transport regret) confirms the arithmetic but is not independent evidence — and the resampled transfer optimum is a pair in at least 97% of draws, which the surrogate interpolates, so the comparison is about fidelity rather than saved runs (11 of 15 coalitions are already fit). The separable decision-level finding is non-identifiability: under the nested bootstrap the top-1 transfer coalition mass spreads across four pairs ($\{\mathrm{agent}, \mathrm{math}\}$ 32.5%, $\{\mathrm{agent}, \mathrm{qa}\}$ 29.1%, $\{\mathrm{agent}, \mathrm{code}\}$ 18.9%, $\{\mathrm{math}, \mathrm{qa}\}$ 16.6%), none past a third, with condition A's seed-0 point estimate (the full set) top-1 in 0.05% of draws. The pairwise surrogate's point-estimate top-1 on transfer does coincide with the full lattice's, but under seed-0 item resampling that agreement survives only 32.3% of draws. Any "best mixture" read off a single-seed lattice of this size is close to a coin toss — which also bounds the 71.5% suite-transport regret rate: with no within-suite control (argmax on a held-out half of P_{tr}), that rate mixes suite shift with estimation noise.

5.4 Curvature: no single Fisher-proportionality constant fits

Proposition 1's literal antecedent is a statement about the reward-weighted H_P^{true} of equation 5. That no source-significant domain-pair dividend flips sign on transfer under condition A (§5.6) is consistent with the antecedent but, proportional curvature being only sufficient for sign agreement, cannot confirm it. We measure the weaker generic condition instead: proportionality of the plain Fisher information F_P each domain singleton's task vector δ_d induces — the domain-generalization Hessian-alignment condition (Hemati et al., 2023) this line of work operationalizes. Sampling 448 items per domain per suite, the decisive statistic is the diagonal ratio $\delta_d^\top F_{\mathrm{src}} \delta_d / \delta_d^\top F_{\mathrm{tr}} \delta_d$ (one c is required for every entry; the diagonals are largest, PSD, and least noisy): agent 0.80 $[0.56, 1.14]$, qa 1.03 $[0.81, 1.30]$, math 2.37 $[2.00, 2.85]$, code 4.64 $[3.64, 5.84]$ — a $5.8\times$ spread, 0.0% of 2,000 item resamples within $1.5\times$ of each other, agent's and code's CIs disjoint. **No single c reproduces this.** A naive through-origin fit reads $\hat{c} = 1.10$, $R^2 = 0.83$ and would misread as support — least squares absorbs exactly this systematic spread — but a scale-free check agrees with the diagonals: normalizing $F[d, d']/\sqrt{F[d, d]F[d', d']}$ cancels c, yet transfer is systematically more curvature-correlated than source in 5 of 6 pairs. What these four domains refute is the reading that would license the absence of sign flips as indirect evidence for the antecedent — one proportionality constant shared across domains on $\mathrm{span}\{\delta_d\}$. A weaker or per-domain alignment condition is not tested, and our own reward-weighted antecedent equation 5 remains unmeasured (Appendix A.3).

5.5 GIFT VS. EXHAUSTIVE ENUMERATION ON THE REAL LATTICE

Against this lattice's exact Möbius transform (B_q of 15 queries):

mean dividend error	B_q=4	B_q=6	B_q=8	B_q=10	B_q=12
$\mathrm{EIG}_{\mathrm{trace}}$ (src / tr)	.013/.012	.010/.010	.007/.007	.006/.005	.004/.003
$\mathrm{EIG}_{\mathrm{logdet}}$ (src / tr)	.016/.015	.016/.015	.015/.015	.014/.015	.011/.011
random, mean (src / tr)	.015/.014	.013/.013	.012/.013	.010/.010	.007/.006

$\mathrm{EIG}_{\mathrm{trace}}$ beats random at every budget on both suites ($10/10$); $\mathrm{EIG}_{\mathrm{logdet}}$ at none ($0/10$), extending the synthetic degeneracy to real data. Replayed on the budget-corrected lattice (correction 1's residuals), $\mathrm{EIG}_{\mathrm{trace}}$ again beats random at every budget ($10/10$) and $\mathrm{EIG}_{\mathrm{logdet}}$ at $1/10$.

5.6 ROBUSTNESS, POWER, AND CONFOUNDS

Condition B — a second algorithm. Clipped GRPO on Qwen2.5-1.5B, same 15 coalitions and frozen suites, own θ_0. **It cannot be budget-corrected** (the sweep exists only for Qwen3-1.7B), so it compares with condition A's *uncorrected* numbers only. Pairs whose interaction dividend flips sign between suites go from **0 (A, and C1) to 5 (B)** of 24; B has more uncorrected order-≥ 3 survivors than A (11/20 tr, 9/20 src) but not its $|\Delta|$-grows-with-order profile (Appendix A.7). B changes base model, algorithm, and a frozen `/no_think` token at once, so neither comparison isolates the algorithm: A's monotone climb and zero sign-flip count are setting-specific, not properties of multi-domain RLVR — consistent with a partly noise-amplified climb (§5.2).

Power. Uncorrected MDE_{80} at order 3 runs 0.037–0.064 on source and 0.028–0.105 on transfer against the 0.028 prior. `tr:code` was both the least-powered suite ($3.7\times$ underpowered, $n = 164$) and the one carrying the largest uncorrected order-≥ 3 dividend — a combination a selection effect predicts as often as a real one, and condition E's order-3 retest did not reproduce it.

Confounds. Seven are registered and carried forward (Appendix A.10): budget dilution (treated above), a `code` memorization pattern a direct test rejects, prompt- rather than FLOP-matched suites, one primary seed, an MDE threshold frozen late, a quadratic-expansion license that fails (touching only Lemma 1's scaffold role), and two baselines.

6 DISCUSSION AND CONCLUSION

The uncorrected headline is not evidence of interaction, and that is the finding. Two claims should be kept apart. The *mechanism* is exact given the single-domain budget-response curves: by equation 3, any coalition-grid audit that scales per-domain training with $|S|$ reports nonzero dividends at every order even when domains do not interact, so raw magnitude-by-order without correction risks the misread this paper started with. The *empirical* corrected null — at most one order-≥ 3 transfer dividend surviving FDR under any route, source unresolved — holds in one setting (GOPO, Qwen3-1.7B, primarily one seed, $|D| = 4$), and our one cross-algorithm check does not reproduce A's order profile or zero sign-flip count (§5.6): the mechanism generalizes, the specific null does not yet. Merging (C1) reproduces the uncorrected source survivors from dilution alone but no transfer ones, so A's transfer dividends are interaction or Join's per-coalition training noise.

What a practitioner should take. Truncating a selection model to pairwise order costs essentially nothing here, whereas choosing on the source suite incurs regret in 71.5% of resamples (Table 2; partly estimation noise, §5.3): the error lives in *which coalition* is picked, not in the dividends, which never significantly flip sign between suites under A or C1 ($0/24$ pairs), and a single-seed grid does not resolve the transfer optimum. To measure interaction, fix the *per-domain* budget.

Limitations. Source is unresolved (6/20 nonparametric, 2/20 nested); σ rests on three replicate pairs; condition E trades the $1/|S|$ confound for a total-compute one and omits `math`; C1 is one seed; Tier B ($|D| = 8$) is unrun; the transfer null is power-limited on `tr:code`/`tr:agent`. A trained four-domain lattice at fixed per-domain budget and a multi-seed run would move it most. The lesson is methodological before empirical: measuring interaction order in a budget-matched grid requires correcting for the grid's own construction first.

AI USE STATEMENT

Generative AI tools were used to search and cross-check related work, review the experiment and analysis code, review the manuscript for internal consistency, check that figures meet publication requirements, and analyse risks to the claims (confounds, statistical power, overclaiming). They also assisted with implementing parts of the analysis code and with editing prose; all such code was reviewed and tested by the authors, and all prose was checked against the underlying result files. No reported number was generated by an AI tool: every result is computed by the code release from actual training and evaluation runs. All theoretical claims were verified by the authors, who take responsibility for all content.

ETHICS STATEMENT

This work audits a statistical property of multi-domain RLVR training mixtures and does not involve human subjects, sensitive personal data, or dual-use capabilities beyond those already present in the cited RLVR training and evaluation suites. We foresee no harms specific to this method beyond those already associated with the underlying models and benchmarks.

REPRODUCIBILITY STATEMENT

The value-function estimand, audit protocol, and decision-relevance/transport-failure definitions (§3–4) are stated precisely enough to be independently implemented; full proofs of Lemma 1 and Proposition 1, and the derivation of equation 5, are in Appendices A.2–A.3. Dataset sources, budgets, and suites are in Table 1. §5's results are reproducible from the code release: `mobius.json` and `budget_inputs_tr.json` back the uncorrected lattice and the closed-form refutation; `budget_null.json` and `run_variance.json` back correction 1; `analysis_E/mobius.json` and `condition_e_compare.json` back correction 2; `nested_bootstrap_{tr,src}.json` and `nested_vcorr.json` back correction 3 and the regret decomposition; `hessian_proportionality.json` backs §5.4; `gift_validation.json` and `gift_validation_corrected.json` back §5.5; `corrected_power.json` backs the corrected power and survivor analysis of Appendix A.6; `condition_b_compare.json` backs §5.6; `analysis_C1{,a}/` and `analysis_C2/` (`mobius.json`, `condition_c_compare.json`, `budget_null{,_A}.json`) back condition C. `scripts/verify_theory.py`'s assertions check every theoretical claim against finite-difference/exact-enumeration ground truth. The pre-registration commit is released alongside the training code.

REFERENCES

Alen Alexanderian. Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: A review. *Inverse Problems*, 37(4):043001, 2021.

Ahmed Attia, Alen Alexanderian, and Arvind K. Saibaba. Goal-oriented optimal design of experiments for large-scale Bayesian linear inverse problems. *Inverse Problems*, 34(9):095009, 2018.

Lukas Balles and Philipp Hennig. Dissecting Adam: The sign, magnitude and variance of stochastic gradients. In *ICML*, 2018.

Landon Butler, Abhineet Agarwal, Justin Singh Kang, Yigit Efe Erginbas, Bin Yu, and Kannan Ramchandran. ProxySPEX: Inference-efficient interpretability via sparse feature interactions in LLMs. In *NeurIPS*, 2025.

Yuen Chen, Haozhe Si, Guojun Zhang, and Han Zhao. Moment alignment: Unifying gradient and Hessian matching for domain generalization. In *UAI*, pp. 705–736, 2025. arXiv:2506.07378.

Zhoujun Cheng, Shibo Hao, Tianyang Liu, Fan Zhou, Yutao Xie, Feng Yao, et al. Revisiting reinforcement learning for LLM reasoning from a cross-domain perspective. *arXiv preprint arXiv:2506.14965*, 2025. NeurIPS 2025.

Steffen Dereich, Arnulf Jentzen, and Sebastian Kassing. ODE approximation for the Adam algorithm: General and overparametrized setting. *arXiv preprint arXiv:2511.04622*, 2025.

Chris Fifty, Ehsan Amid, Zhe Zhao, Tianhe Yu, Rohan Anil, and Chelsea Finn. Efficiently identifying task groupings for multi-task learning. In *NeurIPS*, 2021.

Fabian Fumagalli, Maximilian Muschalik, Patrick Kolpaczki, Eyke Hüllermeier, and Barbara Hammer. SHAP-IQ: Unified approximation of any-order Shapley interactions. In *NeurIPS*, 2023.

Fabian Fumagalli, Maximilian Muschalik, Patrick Kolpaczki, Eyke Hüllermeier, and Barbara Hammer. KernelSHAP-IQ: Weighted least square optimization for Shapley interactions. In *ICML*, 2024.

Fabian Fumagalli, R. Teal Witter, and Christopher Musco. PolySHAP: Extending KernelSHAP with interaction-informed polynomial regression. In *ICLR*, 2026. arXiv:2601.18608.

Amirata Ghorbani and James Zou. Data Shapley: Equitable valuation of data for machine learning. In *ICML*, 2019.

Kimia Hamidieh, Lester Mackey, and David Alvarez-Melis. Domain-aware scaling laws uncover data synergy. *arXiv preprint arXiv:2607.11052*, 2026.

Peter L. Hammer and Sergiu Rudeanu. *Boolean Methods in Operations Research and Related Areas*. Springer, 1968.

Sobhan Hemati, Guojun Zhang, Amir Estiri, and Xi Chen. Understanding Hessian alignment for domain generalization. In *ICCV*, 2023. arXiv:2308.11778.

Chuxuan Hu, Yuxuan Zhu, Antony Kellermann, Caleb Biddulph, Suppakit Waiwitlikhit, Jason Benn, and Daniel Kang. Breaking barriers: Do reinforcement post training gains transfer to unseen domains? *arXiv preprint arXiv:2506.19733*, 2026. ICLR 2026.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. In *ICLR*, 2023. arXiv:2212.04089.

Andrew Ilyas, Sung Min Park, Logan Engstrom, Guillaume Leclerc, and Aleksander Madry. Datamodels: Understanding predictions with data and data with predictions. In *ICML*, 2022. arXiv:2202.00622.

Kimi Team. Kimi K2.5: Visual agentic intelligence. *arXiv preprint arXiv:2602.02276*, 2026.

Patrick Kolpaczki, Maximilian Muschalik, Fabian Fumagalli, Barbara Hammer, and Eyke Hüllermeier. SVARM-IQ: Efficient approximation of any-order Shapley interactions through stratification. In *AISTATS*, 2024.

Andreas Krause and Carlos Guestrin. Near-optimal nonmyopic value of information in graphical models. In *UAI*, 2005.

Dongyue Li, Haotian Ju, Aneesh Sharma, and Hongyang R. Zhang. Boosting multitask learning on graphs through higher-order task affinities. In *KDD*, 2023. arXiv:2306.14009.

Yu Li, Zhuoshi Pan, Honglin Lin, Mengyuan Sun, Conghui He, and Lijun Wu. Can one domain help others? a data-centric study on multi-domain reasoning via reinforcement learning. *arXiv preprint arXiv:2507.17512*, 2025.

Yiqing Liang, Jielin Qiu, Wenhao Ding, Zuxin Liu, James Tompkin, Mengdi Xu, Mengzhou Xia, Zhengzhong Tu, Laixi Shi, and Jiacheng Zhu. MoDoMoDo: Multi-domain data mixtures for multimodal LLM reinforcement learning. *arXiv preprint arXiv:2505.24871*, 2025.

Jinkun Lin, Anqi Zhang, Mathias Lécuyer, Jinyang Li, Aurojit Panda, and Siddhartha Sen. Measuring the effect of training data on deep learning predictions via randomized experiments. In *ICML*, 2022.

Wenhan Ma, Jianyu Wei, Liang Zhao, Hailin Zhang, Bangjun Xiao, Lei Li, et al. MOPD: Multi-teacher on-policy distillation for capability integration in LLM post-training. *arXiv preprint arXiv:2606.30406*, 2026.

S. Aaron McClendon. When model merging rivals joint multi-task reinforcement learning: A task-vector geometry analysis. *arXiv preprint arXiv:2607.16062*, 2026.

Alexandru Meterez, Pranav Ajit Nair, Depen Morwani, Cengiz Pehlevan, Sham Kakade, and Alex Damian. A defense of the quadratic model. *arXiv preprint arXiv:2607.21716*, 2026.

Guillermo Ortiz-Jimenez, Alessandro Favero, and Pascal Frossard. Task arithmetic in the tangent space: Improved editing of pre-trained models. In *NeurIPS*, 2023.

Guillermo Owen. Multilinear extensions of games. *Management Science*, 18(5):P64–P79, 1972.

Barak A. Pearlmutter. Fast exact multiplication by the Hessian. *Neural Computation*, 6(1):147–160, 1994.

Alexandre Ramé, Corentin Dancette, and Matthieu Cord. Fishr: Invariant gradient variances for out-of-distribution generalization. In *ICML*, 2022.

Gian-Carlo Rota. On the foundations of combinatorial theory I: Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2(4):340–368, 1964.

Alexey Rukhovich, Alexander Podolskiy, and Irina Piontkovskaya. Commute your domains: Trajectory optimality criterion for multi-domain learning. In *NeurIPS 2024 Workshop on Mathematics of Modern Machine Learning (M3L)*, 2024. arXiv:2501.15556.

David Rundel, Fabian Fumagalli, Maximilian Muschalik, Bernd Bischl, and Matthias Feurer. ShaplEIG: Bayesian experimental design for Shapley value estimation. *arXiv preprint arXiv:2606.02247*, 2026. ICML 2026.

John Schulman, Sergey Levine, Philipp Moritz, Michael Jordan, and Pieter Abbeel. Trust region policy optimization. In *ICML*, pp. 1889–1897, 2015.

Yuge Shi, Jeffrey Seely, Philip H. S. Torr, N. Siddharth, Awni Hannun, Nicolas Usunier, and Gabriel Synnaeve. Gradient matching for domain generalization. In *ICLR*, 2022.

Trevor Standley, Amir Zamir, Dawn Chen, Leonidas Guibas, Jitendra Malik, and Silvio Savarese. Which tasks should be learned together in multi-task learning? In *ICML*, 2020.

Mukund Sundararajan, Kedar Dhamdhere, and Ashish Agarwal. The Shapley Taylor interaction index. In *ICML*, 2020.

Haoqing Wang, Xiang Long, Ziheng Li, Yilong Xu, Tingguang Li, and Yehui Tang. To mix or to merge: Toward multi-domain reinforcement learning for large language models. *arXiv preprint arXiv:2602.12566*, 2026.

Zixian Wang. Group orthogonalized policy optimization: Group policy optimization as orthogonal projection in Hilbert space. *arXiv preprint arXiv:2602.21269*, 2026.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992.

Yongjin Yang, Jiarui Liu, Yinghui He, Lechen Zhang, Bernhard Schölkopf, and Zhijing Jin. Transferability for general reasoning: An automated curriculum for multi-domain RLVR. *arXiv preprint arXiv:2606.25178*, 2026.

Zhenshuo Zhang, Minxuan Duan, and Hongyang R. Zhang. Efficient estimation of kernel surrogate models for task attribution. *arXiv preprint arXiv:2602.03783*, 2026. ICLR 2026.

A Appendix

A.1 Additional figures

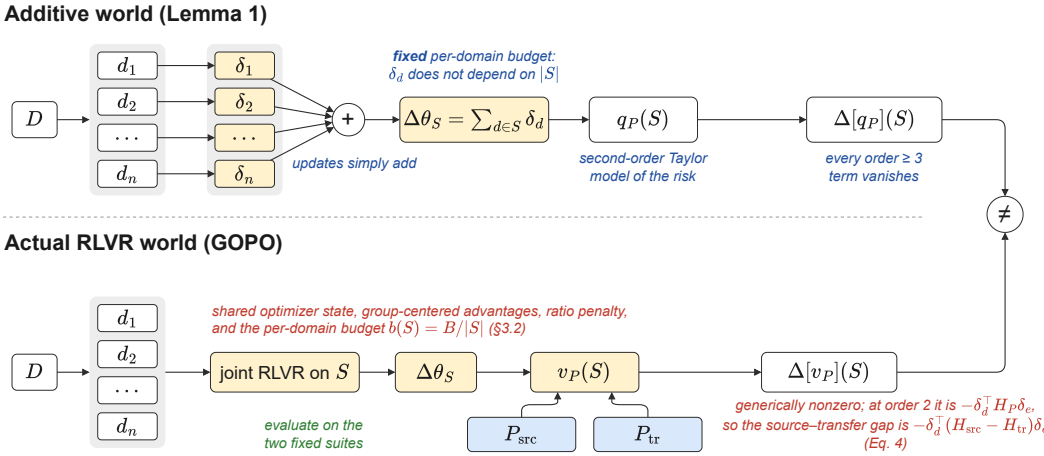


Figure 2: Additive world vs. actual RLVR world. *Top:* at a fixed per-domain budget, δ_d does not depend on $|S|$, updates add, and Lemma 1 forces every order-≥ 3 dividend of the quadratic model q_P to vanish exactly. *Bottom:* actual GOPO training shares optimizer state, centres advantages within a group, penalizes the policy ratio, and divides the budget as $b(S) = B/|S|$, so neither precondition holds; $v_P(S)$ is measured directly on the two frozen suites (blue), and at order 2 the local quadratic model reads $-\delta_d^\top H_P \delta_e$, making the source-transfer gap $-\delta_d^\top (H_{\text{src}} - H_{\text{tr}})\delta_e$ (equation 4, Proposition 1). Yellow marks trained or measured objects, blue the two fixed suites, white derived quantities.

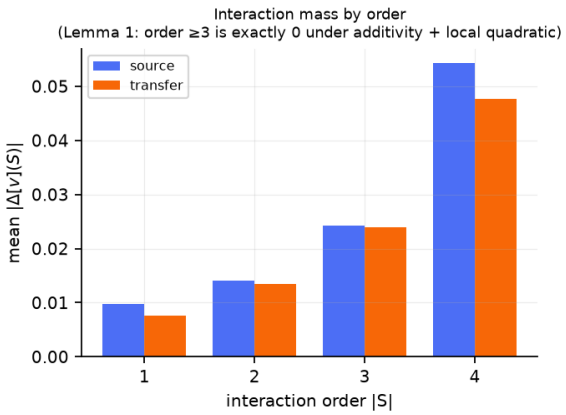


Figure 3: Möbius dividend magnitude by order, source and transfer bars grouped by order, seed-0 lattice. Order 4 is largest on both suites by raw magnitude; the FDR correction of §5.2, not this ordering, is what separates signal from the expected behavior of 20 order-≥ 3 tests under the null, and the pure-noise floor itself rises $3.9\times$ from order 1 to 4.

A.2 Proofs of Lemma 1 and Proposition 1

Quadratic-expansion licensing. The evidence for a quadratic local expansion's predictiveness (Meterez et al., 2026) is gathered for local re-expansion at successive checkpoints, not one expansion spanning a whole run; we would use it only under the additional condition that RLVR post-training compute is comparable to that re-expansion window, so θ_0 plays the role of one re-expansion checkpoint. §5.6(vi) and Appendix A.10 report that this condition, and the weight-space linearity/direction-stability gate, both fail; Lemma 1 is therefore used only as a diagnostic scaffold.

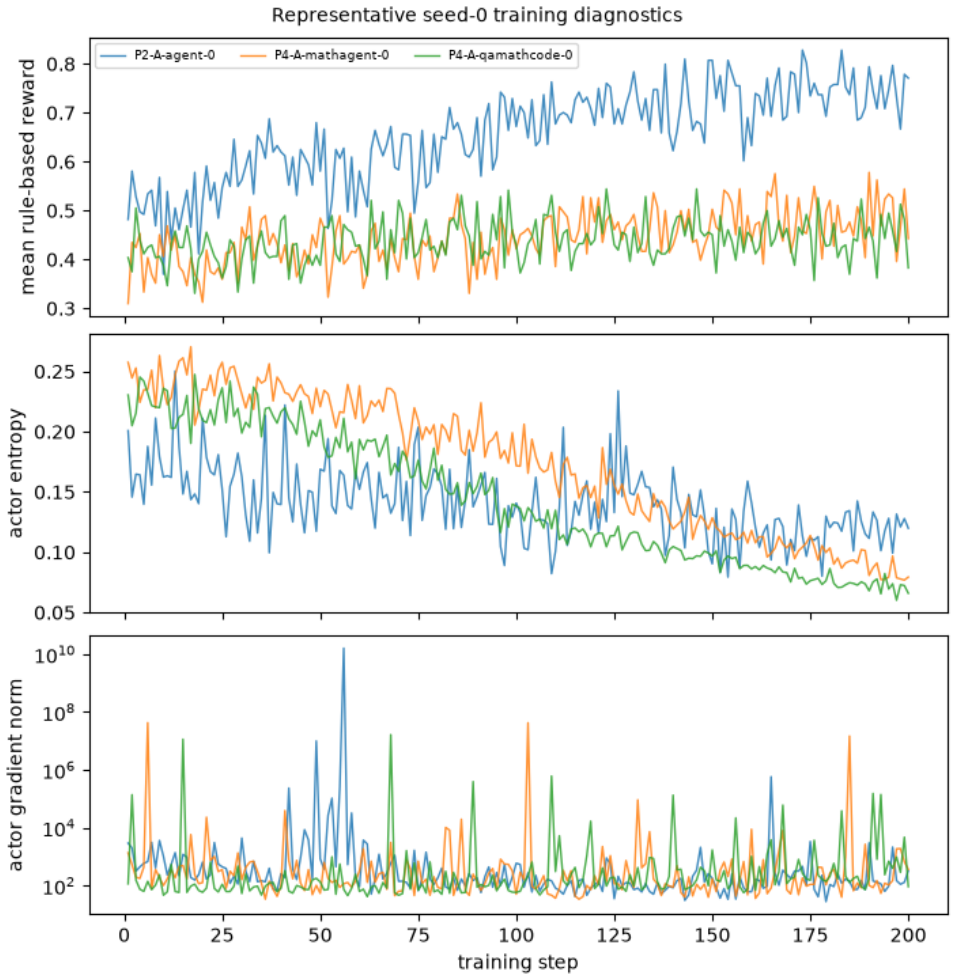


Figure 4: Seed-0 training curves, three representative coalitions: reward rises and stabilizes, entropy decreases smoothly, gradient-norm spikes are transient. A sanity check, not a quantitative result.

Proof. Under additivity, $\Delta\theta_S = \sum_{d\in S} \delta_d$ is a degree-1 multilinear polynomial in the coalition indicator $x \in \{0,1\}^{|D|}$. The quadratic model $q_P(S) = -g_P^\top \Delta\theta_S - \frac{1}{2}\Delta\theta_S^\top H_P \Delta\theta_S$ is then degree ≤ 2 in x (using $x_d^2 = x_d$). By the Hammer–Rudeanu/Owen correspondence between a pseudo-Boolean function's unique multilinear representation and its Möbius transform, a degree-≤ 2 polynomial has $\Delta[\cdot](S) = 0$ for all $|S| \ge 3$. □

Proof of Proposition 1. Under the same model, the order-2 dividend of q_P over $\{d, d'\}$ is $q_P(\{d,d'\}) - q_P(\{d\}) - q_P(\{d'\}) + q_P(\emptyset)$. The linear term cancels, and the quadratic term leaves $-\frac{1}{2}\left[(\delta_d + \delta_{d'})^\top H_P(\delta_d + \delta_{d'}) - \delta_d^\top H_P \delta_d - \delta_{d'}^\top H_P \delta_{d'}\right] = -\delta_d^\top H_P \delta_{d'}$ for symmetric H_P. If $H_{\text{src}} = c\, H_{\text{tr}}$ on $\text{span}\{\delta_d\}$, then $\delta_d^\top H_{\text{src}} \delta_{d'} = c\, \delta_d^\top H_{\text{tr}} \delta_{d'}$ for every pair, so $\Delta[q_{\text{src}}](\{d,d'\}) = c\,\Delta[q_{\text{tr}}](\{d,d'\})$, and $c > 0$ gives the two the same sign (or both zero). □

Derivation of equation 3. Let $|T| = k$ and $g_d(j) = \varphi_d(B/j)$. Since $\tilde{v}_P(U) = \sum_{d\in U} g_d(|U|)$, exchanging the two sums in the Möbius transform gives

$$\Delta[\tilde{v}_P](T) = \sum_{U\subseteq T} (-1)^{k-|U|} \sum_{d\in U} g_d(|U|) = \sum_{d\in T} \sum_{j=1}^{k} \binom{k-1}{j-1} (-1)^{k-j} g_d(j),$$

because the subsets of T of size j that contain a fixed d number $\binom{k-1}{j-1}$. With $j = 1+i$ and $m = k-1$ the inner sum is $\sum_{i=0}^{m} \binom{m}{i}(-1)^{m-i} g_d(1+i) = \Delta^m g_d(1)$, the m-th forward difference. For a linear response $\varphi_d(b) = \alpha_d b$, $g_d(j) = \alpha_d B/j$ and $\sum_{i=0}^{m} \binom{m}{i}(-1)^i/(1+i) = 1/(m+1)$, so $\Delta^{k-1} g_d(1) = (-1)^{k-1}\alpha_d B/k$ and $\Delta[\tilde{v}_P](T) = (-1)^{k-1} B\,\bar{\alpha}_T$, as stated in §3.2. □

Interpreting a measured deviation. Our protocol violates the additivity precondition by construction (§3.2), so an uncorrected residual is uninformative about optimization-level non-additivity; the correction must first remove the budget confound (§5.2). Subject to that, the additivity assumption is the *weight disentanglement* property of the task-arithmetic literature (Ilharco et al., 2023; Ortiz-Jimenez et al., 2023), whose central finding is that disentanglement does *not* reduce to a clean linear/NTK sufficient condition — it is an empirical property of pretraining. A concurrent RL-specific study finds merged and jointly-trained policies become indistinguishable exactly when domain task vectors are near-orthogonal (McClendon, 2026); our measured condition-A cosines range 0.0085–0.0768, the same regime. Condition C tests this directly (Appendix A.8): at matched total budget merged and joint values rank similarly on transfer, but the registered rank-agreement prediction for full-budget experts (C2) fails.

Decision-change / null-rejection dependency. With $|D| = 4$, a degree-≤ 2 pseudo-Boolean model has 11 coefficients fit on the 11 coalitions of size ≤ 2, so it interpolates them exactly and $v^{\text{full}}(S) - \hat{v}_2(S) = \sum_{|T|\geq 3,\, T\subseteq S} \Delta[v](T)$. A decision change (Definition 1) therefore cannot occur unless some order-≥ 3 dividend is nonzero — a free, pre-registered falsification check: a measured decision change while the order-≥ 3 null is not rejected indicates a noise-dominated measurement. This is also why §5.3's order-truncation regret is ≈ 0: it is the decision-level image of §5.2's null, not an independent result.

A.3 From the GOPO surrogate Hessian to the true evaluation Hessian

This appendix derives equation 5.

Scope. The closed form is a modeling substitution, not a claim about clipped GRPO's literal Hessian, though GOPO is an independently validated RLVR algorithm (Wang, 2026). The decomposition relies on GOPO's advantage being mean-only, $A(y) = r(y) - \bar{r}$ (verified against the algorithm's definition, not assumed); a std-normalized advantage as in vanilla GRPO would add a group-reward-variance factor and is not covered. The mechanism is only directly testable when the trained policies are themselves optimized with GOPO, which is why §5 fixes GOPO as the primary condition.

Writing $\pi_\theta(y) = \pi_{\theta_0}(y)\rho_\theta(y)$, the true objective $J(\theta) = \mathbb{E}_{y\sim\pi_{\theta_0}}[r(y)\rho_\theta(y)]$ has, since $r(y)$ does not depend on θ,

$$\nabla^2_\theta J(\theta_0) = \mathbb{E}_{x\sim P,\, y\sim\pi_{\theta_0}}\big[r(y)\big(S(y)S(y)^\top + \nabla^2_\theta \log \pi_{\theta_0}(y)\big)\big], \tag{6}$$

reusing $\nabla^2_\theta \rho_\theta(y)|_{\theta_0} = S(y)S(y)^\top + \nabla^2_\theta \log \pi_{\theta_0}(y)$. Since $H_P^{\text{true}}(\theta_0) = -\nabla^2_\theta J(\theta_0)$, GOPO's surrogate Hessian is $H_P^{\text{gopo}} = \mu F_P - \mathbb{E}[A\,(SS^\top + \nabla^2 \log \pi_{\theta_0})]$; substituting the mean-only advantage $A = r - \bar{r}$ into that expectation and using the information-matrix equality $\mathbb{E}_{y\sim\pi_{\theta_0}}[\nabla^2_\theta \log \pi_{\theta_0}(y)] = -F_P(\theta_0)$ (from differentiating $\sum_y \pi_\theta(y) = 1$ twice), the $\bar{r}$ term vanishes and equation 5 follows. When the empirical group mean $\bar{r} = \frac{1}{G}\sum_i r(y_i)$ is used, the i.i.d. rollouts give $-\mathbb{E}[A\,M] = (1 - \frac{1}{G})H_P^{\text{true}}$ with $M = SS^\top + \nabla^2 \log \pi_{\theta_0}$, so exactly

$$H_P^{\text{true}}(\theta_0) = \frac{G}{G-1}\big(H_P^{\text{gopo}}(\theta_0) - \mu F_P(\theta_0)\big), \tag{7}$$

of which equation 5 is the $G \to \infty$ limit. This is the Hessian-level analogue of the classical result that policy gradients are unbiased under any action-independent baseline (Williams, 1992). The $G/(G-1)$ factor rescales every pairwise dividend in equation 4 by the same scalar and changes no sign or ranking. `scripts/verify_theory.py` verifies equation 7 by exact enumeration over all K^G rollout groups of a categorical policy; residual at machine precision for $G \in \{2, 4, 6\}$.

What equation 5 is and is not tested against. §5.4 measures proportionality of F_P (the reward-independent term), the generic condition (Hemati et al., 2023) operationalizes. It does *not* measure proportionality of H_P^{true} itself, which the same Hessian-vector-product machinery (Appendix A.6) could reach in future work by substituting H_P^{true} for F_P at the same per-sample cost. We do not claim to have run that version. A secondary reading of the same measurement: normalized Fisher-metric correlations reach 0.55 (some pairs are near zero or negative) — an order of magnitude above raw weight-space cosines at that upper end, so near-orthogonality in weight space can still understate the interaction the dividends measure.

A.4 COALITION AUDIT PROTOCOL, STEP BY STEP

Input: D, fixed $P_{\mathrm{src}}, P_{\mathrm{tr}}$, budget B.

1. If $|D| \leq 5$: enumerate every non-empty $S \subseteq D$. Else: use GIFT to select B_q coalitions adaptively.
2. For each S: train π_S by Join with compute split proportionally to $|S|$; measure $v_{\mathrm{src}}(S), v_{\mathrm{tr}}(S)$ with paired eval-item bootstrap CIs. Realized condition A: seed 0 primary; seeds $\{0, 1, 2\}$ for the nested bootstrap only.
3. Compute $\Delta[v_P](S)$ for every S. Correct the $1/|S|$ budget-dilution confound — realized as a nonparametric subtraction, an independent from-scratch design, and a nested seed×item bootstrap (§5.2), not the closed-form q^{budget}, which this paper's own displacements fail to license (Appendix A.6).
4. Apply Lemma 1's null test at $|S| \geq 3$ and the decision-relevance criterion (Definition 1), fixed before any run.

Pre-registration. Lemma 1's significance threshold and the transport taxonomy (within which sign-flip is the primary statistic) are fixed in a timestamped commit before condition A. The decision-relevance minimum effect size and the MDE threshold it depends on were *not*, reported as a pre-registration-discipline miss (§5.6(v)) rather than backfilled. A fully null result was pre-committed as informative on its own, corroborating current low-order practice at the scale tested.

Power design. $\Delta[v](S)$ is a signed sum of $2^{|S|}$ values; if each carries independent error σ, the order-k dividend has standard deviation $\sqrt{2^k - 1}\,\sigma$ — a factor $1.73/2.65/3.87$ at orders $2/3/4$. Detecting a dividend of size ε at 80% power needs $\sigma \leq \varepsilon/(2.8\sqrt{2^k - 1})$. The pre-registered design weighted seeds by lattice position (more seeds for singletons, whose error propagates into every containing dividend); this was retired for the one-seed policy before condition A, and run-to-run training noise at fixed seed is instead measured from accidental replicate pairs (§5.2).

A.5 GIFT: DEFAULT PRIOR AND THE GREEDY GUARANTEE

Multi-target acquisition. Every dividend of interest is a linear functional of v, so the vector of all m target dividends is jointly Gaussian under the GP with covariance $\Sigma_\Delta = W \Sigma_v W^\top$ for the fixed Möbius weight matrix W. GP posterior covariance after a new observation does not depend on the observed value, so the expected information gain of querying S^* reduces to a deterministic change in $\log\det \Sigma_\Delta$ (or $\mathrm{tr}\,\Sigma_\Delta$), computable without Monte Carlo. A fixed-order dividend $\Delta[v](T)$ has support on only the $\leq 2^{|T|}$ subsets of T, so no elementary-symmetric-polynomial reduction (as ShaplEIG needs for the Shapley value) is required. We claim novelty only in identifying the target family and its weight matrix, not in the design criteria (Attia et al., 2018; Alexanderian, 2021).

The guarantee, and its gap. Krause & Guestrin (2005) prove that greedy maximization of mutual information $I(U; A) = H(U) - H(U \mid A)$ — which for jointly Gaussian variables *is* the log-det-reduction criterion — is within $(1 - 1/e)$ of optimal, *provided the candidate variables are conditionally independent given* U. A general GP covariance over the coalition lattice does not satisfy this. Our default prior does, by construction: place independent priors on the target dividends, $\Delta[v](T) \sim \mathcal{N}(0, \tau_T^2)$ for $|T| \leq k + 1$, and set $v(S) = \sum_{T \subseteq S, |T| \leq k+1} \Delta[v](T) + \epsilon_S$ with ϵ_S i.i.d. Gaussian. Then $\mathrm{Cov}(v(S_1), v(S_2) \mid U) = \sigma_\epsilon^2 \delta_{S_1 S_2}$, so the guarantee applies — to $\mathrm{EIG}_{\mathrm{logdet}}$ specifically. $\mathrm{EIG}_{\mathrm{trace}}$ (the objective actually used, and the only one that beats random) has its own OED literature but we have not verified a directly applicable near-submodularity bound and do not claim one — an honest gap. Numerically at $|D| = 3$: the conditional covariance is diagonal to machine precision and greedy selection achieves 0.961 of the brute-force optimum, above the 0.632 bound.

Order ceiling. GIFT estimates dividends up to order $k + 1$ and cannot detect structure above it. In condition A ($|D| = 4$, one order-4 coalition), GIFT's validation against exhaustive enumeration is validation up to order $k + 1 = 3$ only: a nonzero order-4 ground-truth dividend is unrecoverable by construction, and any gap there reflects the prior's ceiling, not the acquisition rule's efficiency. The smooth permutation-invariant kernel alternative removes the ceiling but forfeits the $(1 - 1/e)$ guarantee.

A.6 BUDGET-MATCHING BIAS: FULL DERIVATION AND WHY THE CLOSED FORM FAILS

The $1/|S|$ scaling law. Near θ_0, GOPO's single-domain loss is locally quadratic with constant curvature (§3.3): $\mathcal{L}_d(\theta) \approx \mathcal{L}_d(\theta_0) + g_d^\top(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)^\top H_d(\theta-\theta_0)$. Under gradient flow $\dot\theta = -\nabla\mathcal{L}_d$, the linear ODE solves to $\delta_d^{(b)} = -H_d^{-1}(I - e^{-H_d b})g_d$; budget-matching drives $b = B/|S| \to 0$ as $|S|$ grows, exactly where the small-b expansion $I - e^{-H_d b} = H_d b + O(b^2)$ is most accurate, giving $\delta_d^{(b)} \approx -b\, g_d$ and $\delta_d^{(B/|S|)} \propto 1/|S|$. Real training uses Adam; Dereich et al. (2025) establish that for a vanishing step-size schedule Adam is an asymptotic pseudo-trajectory of a sign-gradient ODE (Balles & Hennig, 2018), giving $\delta_d^{(b)} \approx -b\,\mathrm{sign}(g_d)$ for small b — same $1/|S|$ scaling, different per-domain constant. The qualitative confound does not depend on the optimizer model. Mechanically, with $f(S) = 1/|S|$, $f(\emptyset) = 0$, and $|D| = 3$: $\Delta[f](\{1,2,3\}) = \frac{1}{3} - 3\cdot\frac{1}{2} + 3\cdot 1 - 0 = \frac{11}{6} \neq 0$, confirming nonzero Möbius coefficients beyond degree one; since this holds for the bare size function, the contamination recurs at every order the lattice supports (equation 2).

Two estimands. Fixing the per-domain budget makes δ_d independent of $|S|$, satisfies Lemma 1's precondition exactly, and contributes zero to every Möbius coefficient of order ≥ 2. We fix total budget nonetheless, because the two protocols answer different questions: constant per-domain budget isolates interaction structure; fixed total budget is the estimand a practitioner faces. We report every $\Delta[v_P](S)$ with and without correction so the two stay separable.

Why the closed-form correction is not applied. Computing q^{budget} needs only g_P (the score-function estimate over rollouts already sampled for $v_P(\emptyset)$) and one Hessian-vector product in the fixed direction $\Delta\theta^{\mathrm{budget}}$, via the R-operator (Pearlmutter, 1994; Schulman et al., 2015) — no full Hessian. But the leading-order expansion is unlicensed at the displacements these runs reach. On a synthetic quadratic with realistic conditioning, the displacement stays linear in b to within 5% and keeps $\cos(\delta_d^{(b)}, -\mathrm{sign}(g_d)) \geq 0.99$ only for $b \leq 200$ steps. The real condition-A singletons exit this window immediately: the pre-registered $R^2 \geq 0.95$ / $\cos \geq 0.9$ gate fails on every domain ($\cos \in [0.10, 0.29]$ against the 0.90 threshold). Measured directly at θ_0 on transfer for the full coalition: $\|\delta\| = 0.214$, $g_P^\top\delta = +2.58$ (a first-order change in a reward bounded in $[0,1]$), $\delta^\top H_P \delta = -0.81$, so $q^{\mathrm{budget}} = -2.18$ against a measured v_{tr} of $+0.0078$ — $279\times$ the quantity it corrects. Three instrumentation defects that had silently zeroed g_P (a missing chat template, a 384-token completion ceiling, gradient checkpointing left on during generation) were found and fixed before this measurement, by inspecting raw completions rather than trusting a zero. Corrections 1–3 in §5.2 are expansion-free and unaffected.

Correction 1, in full. The budget sweep (12/12 runs) trains each domain alone at $b \in \{0.25, 0.5, 0.75, 1.0\}B$ and evaluates on every benchmark, giving the single-domain score gain $\varphi_d^P(b)$. Since $\{1, 0.5, 0.25\}B$ are what $|S| = 1, 2, 4$ realize, the additive null $\tilde v_P(S) = \sum_{d\in S}\varphi_d^P(B/|S|)$ is computable at every size except $|S| = 3$ ($B/3$, interpolated). Built only from single-domain runs, so the test against the full lattice is not circular: it asserts cross-domain additivity at the budget each domain received. Order-1 residual dividends are identically zero by construction (verified before trusting any higher-order residual). Möbius dividends of $v_P(S) - \tilde v_P(S)$, BH-FDR at $q = 0.1$ over the same 60-test-per-suite family: transfer $6/20 \to 0/20$, source $8/20 \to 6/20$; identical surviving set under linear and log interpolation. Folding the measured run-to-run $\sigma \approx 0.004$–0.005 (from three replicate pairs) into the FDR test as an added-quadrature sensitivity leaves transfer at $0/20$ and moves source to $4/20$ — the source shift illustrates the correction's sensitivity on the noisier suite, not a second independently trustworthy count. A dependence-robust Benjamini–Yekutieli re-run of the raw table lowers transfer's order-≥ 3 count from 6 to 2 and leaves source's unchanged at 8, consistent with the corrected conclusion.

Power after correction 1. The registered MDE covers the uncorrected dividends; for the corrected ones we apply the same formula ($\mathrm{MDE}_{80} = 2.80\times$ the mean standard error over a benchmark's order-≥ 3 targets), with each standard error read off its 95% bootstrap CI. The same CI-to-SE conversion applied to the uncorrected dividends reproduces their recorded standard errors to a ratio of 1.00 (range 0.97–1.03). Corrected MDE_{80} is 0.036–0.095 on transfer and 0.048–0.063 on source, 0.91–$1.34\times$ the uncorrected values, so subtracting the null does not blunt the test. It also does not remove every former survivor. Of the six transfer dividends significant before correction, the four three-ways shrink to 40–77% of their size and lose significance; the two order-4 dividends keep their size (`tr:code` $+0.163 \to +0.136$, CI $[+0.041, +0.234]$; `tr:math` $+0.038 \to +0.041$, CI

$[+0.005, +0.075]$) and fall below significance only under BH-FDR across the 60-test family. Five of the 20 corrected transfer dividends have CIs excluding zero before that step. Both the corrected and the uncorrected tables use $1{,}000$ bootstrap draws, which floors p-values at 10^{-3}, close to the BH rank-1 threshold $0.1/60 \approx 1.7 \times 10^{-3}$; the transfer $0/20$ should be read with that resolution in mind.

Correction 1 with $B/3$ measured. The registered analysis interpolates $\varphi_d(B/3)$; the four $b = B/3$ sweep runs were completed after it was computed, and the reported correction was not recomputed. Recomputed on the same condition-A lattice with $B/3$ measured (so the two interpolants coincide), transfer goes $6/20 \to 1/20$ and source $8/20 \to 2/20$. The transfer survivor is the order-4 `tr:code`|{agent,code,math,qa} dividend above, $+0.163 \to +0.211$, CI $[+0.080, +0.347]$; the source survivors are `src:math` and `src:qa` on {agent,code,qa} ($-0.088 \to -0.103$, $-0.055 \to -0.093$). The order-4 `tr:code` term is therefore the one transfer dividend whose survival depends on how $B/3$ is filled; it sits on the least-powered benchmark (§5.6) and is reported, not resolved. Artifact: `analysis_C1/budget_null_A.json`.

Correction 2 (condition E), in full. Each member domain of $\{\text{agent}, \text{code}, \text{qa}\}$ receives the full per-domain budget $B/2$ regardless of coalition size, so $|S|$ never divides the budget. Cost: a one-directional total-compute confound (larger coalitions train more total steps). Prediction committed in the pre-registration record before any of its seven runs completed. On the three benchmarks common to both conditions for $\{\text{agent}, \text{code}, \text{qa}\}$: transfer `tr:code` is FDR-significant under A (-0.093) and not under E (-0.032), and E's MDE_{80} there (0.063) is below A's effect, so not a power shortfall; source `src:qa` (-0.055 under A) becomes non-significant and ***flips sign*** under E ($+0.005$) with disjoint CIs. E instead flags `src:agent` and `src:code`, which A did not. A repeated-domain control (one domain trained alone at E's 100/200/300-step schedule) shows 0 of 5 E dividends surviving FDR on this coalition exceed the mean compute-and-schedule artifact ceiling, though 2 of 5 — both of E's new source significances — exceed a more conservative (replicate-max) version of that ceiling, so those two are not cleanly separated from E's own schedule confound. Two independent methods (this lattice and correction 1, sharing no sweep data) agree that A's transfer order-≥ 3 signal does not survive removing the confound, and that A's source signal does not reproduce on this sub-lattice. Condition C1 (Appendix A.8) adds that dilution alone produces no transfer signal, so which mechanism generated A's is left open.

Budget-sweep memorization numbers. On `code` and `math`, $|v_{\text{tr}}|$ shrinks monotonically as budget falls (`code`: $0.0111 \to 0.0034$; `math`: $0.0077 \to 0.0043$), consistent with memorization; `qa` and `agent` are non-monotone. A half-corpus control run separates repetition from training amount: 237 problems (half the corpus) at the same 19,200-prompt budget, ≈ 81 epochs vs. ≈ 40. Observed: `tr:code` improves ($-0.052 \to -0.038$, 2.6σ) and v_{src} improves (3.9σ) — the *opposite* sign to the memorization prediction committed before the run. The amount of training, not repetition, drives the code domain's transfer failure.

A.7 Nested seed-and-item bootstrap and Condition B

Seed provenance. Seed 2 completed 15/15; seed 1's provenance was re-audited against the cutoff commit for the single-draw-evaluation fix (15/15 pass), so seeds $\{0, 1, 2\}$ are poolable. The nested bootstrap resamples seeds and items jointly over 10 exact seed multisets; every earlier order-≥ 3 number used an item-only bootstrap, blind to seed-to-seed training variance by construction.

Order-≥ 3 under the nested bootstrap. Transfer $6/20 \to 1/20$, source $8/20 \to 2/20$. The three survivors, all `math`:

benchmark	coalition	Δ	CI_{95}
`tr:math`	agent, math, qa	-0.0199	$[-0.0327, -0.0062]$
`src:math`	agent, math, qa	-0.0427	$[-0.0675, -0.0167]$
`src:math`	agent, code, math	-0.0335	$[-0.0501, -0.0170]$

This is the third route (correction 1: nonparametric subtraction, transfer $0/20$; correction 1 + measured σ sensitivity: transfer $0/20$, source $4/20$; this: transfer $1/20$, source $2/20$). The routes share the seed-0 condition-A lattice and both additivity-based routes take cross-domain additivity as the

null; what differs is the facet each removes — systematic budget bias via measured response curves, design-level removal, and seed-variance accounting. They converge on the transfer suite.

Regret decomposition. Resampling Definition 1's three regrets over the same seed×item draws (1998 draws): order-truncation nonzero in 0.05% (median 0.0000, mean 3×10^{-10}); suite-transport nonzero in 71.5% (median 0.0025, $p_{97.5} = 0.0229$); compound nonzero in 68.7% (median 0.0019). Top-1 transfer coalition frequency: $\{\text{agent}, \text{math}\}$ 32.5%, $\{\text{agent}, \text{qa}\}$ 29.1%, $\{\text{agent}, \text{code}\}$ 18.9%, $\{\text{math}, \text{qa}\}$ 16.6%, remainder $< 3\%$ each, full four-domain set 0.05%.

Condition B. Clipped GRPO on Qwen2.5-1.5B, same 15 coalitions and frozen suites, own θ_0 ($P_{\text{src}} = 0.4126$, $P_{\text{tr}} = 0.4285$). B is *not* budget-corrected and cannot be (the nonparametric correction needs a per-domain sweep that exists only for Qwen3-1.7B), so B's counts are comparable only to condition A's uncorrected numbers. Individual $v_P(S)$ are not comparable across conditions (different base model, and a `/no_think` token frozen into every prompt for Qwen3's benefit that is noise to Qwen2.5). Two scale-free comparisons are valid across models: the sign-flip count over the 24 pairwise dividends (6 domain pairs $\times$ 4 benchmark scores) goes 0 (A) to 5 (B); the $|\Delta|$-by-order profile, normalized to each lattice's order-1 scale, is $1.00/1.79/3.18/6.31$ (A, transfer) vs. $1.00/0.33/0.56/1.49$ (B, transfer), shape correlation 0.67; yet B has more uncorrected order-≥ 3 FDR survivors than A (11/20 tr and 9/20 src, against 6/20 and 8/20). The order-truncation top-1 (pairwise surrogate vs. full lattice, transfer) does not flip under B either, but that conclusion survives only 35.9% of seed-0 item resamples (A: 32.3%). Artifacts: `results/analysis_B/mobius.json`, `condition_b_compare.json`.

A.8 Condition C (Distill): lattices built by merging

Construction. $\theta_S = \theta_0 + \lambda \sum_{d \in S} \delta_d(b)$, with $\delta_d(b) = \theta_d(b) - \theta_0$ the task vector of domain d trained alone (GOPO, condition A's recipe) at per-domain budget b. No coalition is trained; the merge is a deterministic tensor-wise sum in float64, and every merged policy is scored by condition A's evaluation code. Three complete lattices (15 coalitions each, singletons included) were registered in the pre-registration record before any merged policy was evaluated, together with a merge-health gate ($v_{\text{src}}(S) < -0.05$ marks a failed merge and voids every dividend touching it) and predictions H1–H5: **C1** (primary): $b = B/|S|$, $\lambda = 1$ — total budget B, so the $1/|S|$ dilution is present exactly as in A; **C1a**: as C1 with $\lambda = 1/|S|$; **C2**: $b = B$, $\lambda = 1$ — total compute grows with $|S|$, as in condition E. The experts are the singleton and the $b \in \{B/2, B/3, B/4\}$ sweep runs, re-run at seed 0 with the same recipe to obtain weights for merging; C1's correction 1 is therefore built from the same expert runs as its lattice, and $B/3$ is measured, not interpolated. Condition A's own lattice and correction are left on the original runs. Pairwise task-vector cosines are ≤ 0.096. Mean $\|\Delta\theta_S\|$ by $|S| = 1..4$ is $0.40/0.37/0.34/0.32$ under C1, $0.40/0.19/0.11/0.08$ under C1a, and $0.40/0.58/0.72/0.84$ under C2: only C1 holds the displacement roughly fixed across sizes.

Table 3: Condition C against condition A. Uncorrected counts: order-≥ 3 dividends surviving BH-FDR ($q = 0.1$, 60 tests per suite), reportable after the merge-health gate. Profile: suite-pooled mean $|\Delta|$ at orders 1–4, normalized to order 1.

	tr	src	tr profile	src profile	failed merges
A (Join)	6/20	8/20	$1/1.79/3.18/6.31$	$1/1.44/2.50/5.59$	—
C1 ($b{=}B/\|S\|$, $\lambda{=}1$)	0/20	11/20	$1/1.19/1.20/0.47$	$1/1.66/1.96/0.84$	none
C1a ($b{=}B/\|S\|$, $\lambda{=}1/\|S\|$)	1/20	8/20	$1/1.26/1.80/2.73$	$1/1.20/2.78/3.71$	none
C2 ($b{=}B$, $\lambda{=}1$)	0/20	2/20	$1/1.36/0.41/1.84$	$1/1.55/2.11/2.81$	{code,math,qa}
C1 after correction 1	1/20	0/20	A with $B/3$ measured: 1/20, 2/20		

H1 (signature without joint training): mixed. Registered prediction: on the uncorrected C1 lattice $|\Delta|$ rises with order and some order-≥ 3 dividend survives FDR; after correction 1 none survives on transfer. On source, 11/20 survive, 6 of them among A's 8, and correction 1 removes all 11. On transfer none survives even uncorrected, and neither suite's profile rises. The falsification clause (no reportable survivor on either suite) is not met, but the transfer half of A's signature is not reproduced. After correction one transfer residual survives, `tr:agent`|{code,math,qa}, $-0.009 \to +0.023$, CI $[+0.007, +0.041]$, a dividend the correction creates rather than one it fails to remove.

What C1 does and does not separate. Under C1 the null $\sum_{d\in S}\varphi_d(B/|S|)$ is built from the very expert runs that are merged, so correction 1 removes their training noise along with the dilution; under Join each coalition is a separate training run whose noise enters its value once, and the noise floor of §5.2 grows with order. Condition A's order-≥ 3 dividends are larger than C1's, and C1 cannot say whether the excess is joint-training interaction or that per-coalition noise; with one seed per lattice it is not separable here.

Size-only terms. C1a multiplies every task vector by $1/|S|$, a coefficient that depends on nothing but coalition size, and its profile rises monotonically on both suites ($2.73\times$ and $3.71\times$ at order 4) where C1's does not. This is the mechanism of §3.2 arriving through the merge coefficient instead of the budget, which is why $\lambda = 1$ was registered as primary.

H2 (attributing correction 3's survivors). Registered rule: present with the same sign under C1 $\Rightarrow$ a property of the domains; absent where C1's MDE_{80} is at most A's effect $\Rightarrow$ attributable to joint training; otherwise inconclusive. The two source survivors are present with the same sign: `src:math`|{agent,math,qa} $-0.043 \rightarrow -0.029$, CI $[-0.052, -0.008]$, and `src:math`|{agent,code,math} $-0.034 \rightarrow -0.023$, CI $[-0.047, -0.001]$, both FDR-significant. The transfer survivor, `tr:math`|{agent,math,qa} (-0.020 under A), is $+0.004$, CI $[-0.014, +0.022]$, under C1, but C1's MDE_{80} there is 0.027: inconclusive.

H3 (Join vs. merge values). Registered for C2 as rank agreement (Spearman > 0.5 per suite), predicted from near-orthogonal task vectors (McClendon, 2026). It is *falsified*: 0.25 on transfer, -0.20 on source, mean $|v_{\mathrm{C2}} - v_{\mathrm{A}}|$ 0.018/0.021. C2's values fall with $|S|$ (full coalition $v_{\mathrm{tr}} = -0.064$): joint training within total budget B yields more than merging experts each trained at B. At matched budget (C1, not registered) agreement is closer: Spearman 0.70 on transfer and 0.37 on source, mean $|v_{\mathrm{C1}} - v_{\mathrm{A}}|$ 0.004/0.008.

H4 (transport). Transport failure holds under C1 as under A: top-1 on source {code,math,qa} vs. {agent,math,qa} on transfer (A: {code,qa} vs. the full coalition), and $0/24$ domain-pair dividends flip sign significantly between suites under A, C1, or C2. The top-1 coalitions themselves differ between constructions.

H5 (non-quadratic residual). Because additivity holds by construction under C1, an order-≥ 3 dividend surviving correction 1 measures the non-quadratic part of R_P along $\mathrm{span}\{\delta_d\}$. One of 40 survives (above), and order-3 MDE_{80} on C1 is 0.027–0.085 (transfer) and 0.035–0.056 (source): no non-quadratic term of that size is present at these displacements, a statement about the achieved power, not a proof of the quadratic model.

C2. Without dilution, C2 gives $0/20$ transfer survivors on all four domains including `math`, which condition E lacks, but it is not a clean null: $\|\Delta\theta_S\|$ doubles from $|S| = 1$ to 4, its values fall with $|S|$, and {code,math,qa} fails the merge-health gate ($v_{\mathrm{src}} = -0.064$). Its two reportable source survivors, the same two `src:math` three-ways as above, carry the opposite sign to A's ($+0.038$, $+0.031$). C2 was run at seed 0 only; the registered seeds 1–2 are not run. Artifacts: `results/analysis_C1{,a}/`, `results/analysis_C2/` (`mobius.json`, `condition_c_compare.json`), and `analysis_C1/budget_null{,_A}.json`.

A.9 Coalition values, all 15 condition-A coalitions

Table 4 lists the raw coalition values behind §5.2. $v_P(S) = \mathrm{score}_P(\theta_S) - \mathrm{score}_P(\theta_0)$; θ_0: src $= 0.5230$, tr $= 0.5709$. *Transport* labels whether $\mathrm{sign}(v_{\mathrm{src}})$ carries to v_{tr} at comparable magnitude (carries), shrinks by $> 40\%$ without flipping (attenuated), or flips (sign flip) — a coalition-level descriptive convenience, distinct from the pre-registered 0.5 attenuation ratio applied at the dividend level. These are point estimates; no coalition-scalar CI is available.

v_{src} is positive for 13/15 coalitions and generally rises with $|S|$; v_{tr} is smaller and oscillates in sign (4 sign flip, 8 attenuate, 3 carry). This is the shape the transport question is about, but coalition-level point estimates and the statistically-resolved dividends of §5.2 count different objects and need not agree on which orders drive it.

Table 4: Coalition values, all 15 condition-A coalitions (seed 0).

coalition	v_{src}	v_{tr}	transport
{agent}	$+0.0156$	$+0.0017$	attenuated
{code}	$+0.0036$	-0.0174	sign flip
{math}	-0.0160	-0.0077	attenuated
{qa}	$+0.0036$	$+0.0034$	carries
{agent,code}	$+0.0081$	$+0.0005$	attenuated
{agent,math}	$+0.0131$	$+0.0076$	attenuated
{agent,qa}	$+0.0165$	$+0.0025$	attenuated
{code,math}	$+0.0056$	$+0.0022$	attenuated
{code,qa}	$+0.0215$	-0.0002	sign flip
{math,qa}	$+0.0122$	$+0.0035$	attenuated
{agent,code,math}	$+0.0205$	-0.0001	sign flip
{agent,code,qa}	-0.0080	-0.0056	carries
{agent,math,qa}	$+0.0110$	$+0.0069$	carries
{code,math,qa}	$+0.0131$	-0.0048	sign flip
{agent,code,math,qa}	$+0.0206$	$+0.0078$	attenuated

A.10 Known confounds and deviations, full detail

Table 1 deviations. GPQA-Diamond is HF-hub-gated; its absence is a documented reduction of $P_{\text{tr}}(\text{qa})$. CodeForces was abandoned before any run (θ_0 0.010 pass rate, $\approx 96\%$ all-zero-reward GRPO groups); CodeElo and LCBv6 dropped, not substituted; MBPP→HumanEval is a mild shift (both short Python function synthesis), weighted accordingly. BFCLv3 MT base gives only 100 training prompts ($\approx$ 192 epochs vs. $\approx$ 10–12 for `math`, `qa` and `agent`); replaced with BFCL's AST-verifiable single-turn categories, a synthetic-to-real source→transfer split.

Seven registered confounds. (i) $1/|S|$ budget-dilution — closed-form correction inapplicable (279× overshoot, Appendix A.6); removed two data-driven ways plus a partial from-scratch retrain (§5.2); removing it eliminates the transfer-suite order-≥ 3 result under the data-driven routes, and shows the source result does not reproduce on the one sub-lattice retested. (ii) `train_code` holds 474 unique problems vs. 1652–2000 elsewhere; the memorization-consistent sweep pattern does not survive the half-corpus test (Appendix A.6), reported descriptively only. (iii) Prompt-matched, not FLOP-matched: 16× response-length range. (iv) One primary training seed; three completed and used only for the nested bootstrap; the exclusion of seed-1/2 from other estimates is content-blind, applied identically to every coalition. The one decision-relevance close call — the order-truncation rule's no-flip conclusion survives only 32.3% of seed-0 eval-item resamples (§5.3) — is reported at this scope rather than resolved by adding seeds. (v) The MDE threshold in `prereg.json` was never frozen before the first result landed; every MDE number is reported against the silent 0.028 fallback, disclosed rather than backfilled. (vi) The quadratic-expansion's FLOP-ratio condition (RLVR post-training vs. pretraining FLOPs against Meterez et al. (2026)'s 5–10% window) is checked and fails by roughly four orders of magnitude (0.0001–0.0006%), alongside the weight-space gate that already fails. All three corrections used in §5.2 are expansion-free, so this touches only Lemma 1's scaffold role. (vii) Two baselines are reported: the uniform-mixture baseline (the already-trained full coalition) and a MoDoMoDo-style continuous quadratic-in-mixture-weight fit against all 15 coalitions (ridge-regularized, $\lambda \approx 1.94 \times 10^{-6}$, condition number $10^{17} \rightarrow 10^{7}$; λ-sensitivity across four orders of magnitude leaves the conclusion unchanged) whose simplex optimum, rounded to the nearest measured coalition, disagrees with the measured top-1 on both suites (transfer: $\{\text{agent}, \text{math}\}$ vs. the full four-domain top-1) — a second line of evidence for §5.3. The TAC-style baseline needs live per-step gradient projections not reconstructable from post-hoc coalition values; no static substitute is offered under that name. The Li-et-al.-retrained baseline and the domain-label-permutation and budget-matched-vs-raw controls are not included.

Per-domain power ratios. MDE_{80} at order 3 against the 0.028 prior: `tr:code` underpowered $3.7\times$ ($n = 164$), `tr:agent` $2.6\times$ ($n = 258$), `tr:math` $1.0\times$, `tr:qa` $1.1\times$. The suites carrying the most source signal (`math`, `qa`) are close to adequately powered on transfer; `tr:code` and `tr:agent` are not. Evaluation sampling is $\leq 12\%$ of per-item variance everywhere, so only more items would help, and the suites are frozen.