DECISION-AWARE SAFE CAUSAL BAYESIAN OPTIMIZATION VIA STRUCTURAL VALUE OF INFORMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Causal Bayesian optimization seeks interventions that improve the decision objective under limited information. When the causal graph is unknown, the interplay among structural uncertainty, intervention effects and execution risk further complicates intervention selection. However, reducing structural uncertainty does not necessarily benefit downstream decision-making; the value of structural information depends on its impact on intervention decisions. In this paper, we propose a decision-aware and safe sequential causal Bayesian optimization framework that explicitly evaluates the value of structural information through its impact on downstream intervention decisions. We define Structural Value of Information (SVI) as the expected reduction in the best risk-penalized loss under a hypothetical graph-posterior update with fixed decision maps, and combine it with predicted objective and intervention cost to form an acquisition criterion. To improve computational efficiency, we perform decision-aware clustering of graph hypotheses based on their induced objective and risk maps, and aggregate each cluster's posterior mass onto a representative decision map. Fresh simulations from the full posterior predictive mixture certify candidate joint violation probabilities. The resulting safety guarantee is model-based; transfer to the physical system requires an explicit bound on model discrepancy. Through synthetic experiments and real-world case studies, we demonstrate the effectiveness of the proposed method in balancing optimization performance, structural exploration and safe execution.

1 INTRODUCTION

Causal Bayesian optimization (CBO) uses causal structure to select interventions when experimentation is costly (Aglietti et al., 2020). Operational constraints require optimizing performance while accounting for adverse outcomes (Aglietti et al., 2023; Eriksson & Poloczek, 2021; Gardner et al., 2014; Nguyen et al., 2024). With an uncertain graph, competing hypotheses can predict different intervention effects and violation risks. The learner must therefore decide both which intervention to execute and which structural uncertainty is worth resolving.

Causal optimization can jointly learn graph structure and improve an intervention target (Branchini et al., 2023). However, reducing graph uncertainty need not improve the decision: a large information gain may leave the preferred intervention unchanged, whereas a small posterior shift may reverse its ranking. Recent work shows that complete structural recovery can be unnecessary for causal decision-making (Elahi et al., 2024; Park et al., 2025), and that graph identification and regret minimization can pursue different objectives (Shahverdikondori et al., 2026). This raises the central question of how to value structural information by its effect on downstream intervention decisions.

We introduce Decision-Aware Safe Causal Bayesian Optimization (DASCBO), centered on Structural Value of Information (SVI). SVI measures the expected reduction in the best risk-penalized loss after a hypothetical graph-posterior update. Holding the current objective and risk maps fixed isolates the value of structural reweighting from mechanism refinement. The acquisition criterion combines SVI with the predicted objective and intervention cost.

To make repeated SVI evaluation tractable, we cluster graphs by their objective and risk maps on the candidate set and aggregate posterior mass onto representative maps. We retain every graph's likelihood contribution because similar decisions need not imply similar observations. Before execution,

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.

fresh simulations from the full posterior predictive mixture certify joint violation probabilities. Our contributions are:

- **Decision-aware structural information.** We define SVI, establish nonnegativity and a mutual-information upper bound, and show that it vanishes when all graph hypotheses share an optimal intervention.
- **Controlled decision-space compression.** We bound the error in compressed SVI through pilot estimation, decision-map compression, and Monte Carlo evaluation while preserving full cluster evidence.
- **Sequential safety certification.** Independent full-mixture certification yields simultaneous posterior-predictive safety guarantees. Explicit model-discrepancy conditions give physical-risk and conditional decision-quality bounds.

2 Background and Related Work

2.1 Problem Formulation

We model the system using a structural causal model (SCM) on endogenous variables $\mathcal{V} = \{V_1, \dots, V_p\}$ (Pearl, 2009; Peters et al., 2017). Its directed acyclic graph G and structural equations

$$V_i = f_i\big(V_{\mathrm{pa}_G(i)}, E_i\big), \qquad i = 1, \dots, p,$$

specify how each variable depends on its parents and exogenous noise, where $E = (E_1, \dots, E_p)$ has joint law P_E. An action x in a compact intervention domain $\mathcal{X}$ specifies admissible target variables and their assigned values. The intervention $do(x)$ replaces the corresponding structural equations by these constants and induces the true interventional distribution $P^\star_x = P^{\mathrm{true}}(\cdot \mid do(x))$.

Let $Y \in \mathcal{V}$ be the objective to minimize and $C_1, \dots, C_J \in \mathcal{V}$ be constraint variables with thresholds $\lambda_1, \dots, \lambda_J$. We measure risk by the probability that any constraint is violated:

$$Z = \mathbf{1}\{\exists j \in \{1, \dots, J\} : C_j > \lambda_j\}, \qquad r^\star(x) = P^\star_x(Z = 1).$$

With $J^\star(x) = \mathbb{E}_{P^\star_x}[Y]$ and joint violation budget $\alpha \in (0, 1)$, the optimization problem is

$$\min_{x \in \mathcal{X}} J^\star(x) \qquad \text{subject to} \qquad r^\star(x) \le \alpha. \tag{1}$$

This joint chance constraint controls the probability of any violation without assuming independence among constraints (Shapiro et al., 2021). The graph, mechanisms, and noise law are unknown. At round t, the observed history $\mathcal{D}_t$ informs the choice of an intervention x_t; its recorded outcome o_t then updates the model. Section 3.1 defines the posterior objective and risk maps, and Section 3.3 gives the certification rule for execution.

2.2 Related Work

Causal optimization under structural uncertainty. CBO uses a known graph to guide intervention selection (Aglietti et al., 2020). Model-based CBO propagates uncertainty through causal mechanisms (Sussex et al., 2023), and constrained CBO incorporates constraints on intervention outcomes (Aglietti et al., 2023). CEO jointly learns graph structure and optimizes the target through information acquisition (Branchini et al., 2023). Graph-agnostic and unknown-graph approaches further relax structural knowledge requirements (Durand et al., 2025; Mukherjee et al., 2024). DASCBO values structural updates by their effect on the best risk-penalized intervention decision.

Decision-aware information acquisition. Causal bandits exploit causal information to identify effective interventions (Lattimore et al., 2016; Lee & Bareinboim, 2018; Lu et al., 2020; Sen et al., 2017). Partial structural recovery can suffice for no-regret learning (Elahi et al., 2024; Lu et al., 2021), while parent identification can conflict with regret minimization (Shahverdikondori et al., 2026). These results concern bandit regret; our focus is constrained Bayesian optimization over uncertain causal models. SVI builds on the decision-theoretic value of information and knowledge-gradient policies (Frazier et al., 2008; 2009; Howard, 1966), specializing them to graph-posterior reweighting with fixed objective and risk maps.

Constrained and safe Bayesian optimization. Constrained BO addresses unknown feasible regions, including scalable search (Eriksson & Poloczek, 2021) and decoupled objective and constraint evaluations (Nguyen et al., 2024). SafeOpt restricts exploration to a high-confidence safe region (Sui et al., 2015), with extensions to multiple safety constraints (Berkenkamp et al., 2023). More recently, online conformal prediction has been used to control the fraction of unsafe evaluations (Zhang et al., 2025). DASCBO instead certifies each selected action's joint violation probability under the full posterior predictive law. Decision-map compression affects acquisition, while physical-risk guarantees additionally require control of model discrepancy.

3 METHODOLOGY

In this section, we present the Decision-Aware Safe Causal Bayesian Optimization (DASCBO) framework for sequential intervention selection under causal-structure uncertainty and safety constraints. DASCBO evaluates structural information by its impact on downstream decisions, compresses graph uncertainty in decision space for efficient acquisition, and certifies interventions using the full posterior predictive mixture. Figure 1 illustrates the overall architecture of DASCBO.

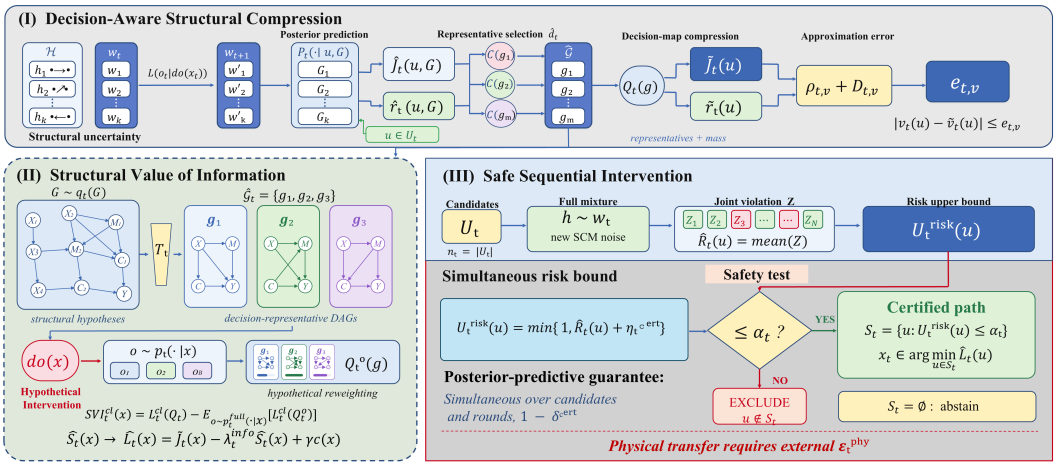


Figure 1: Overview of the DASCBO framework.

3.1 DECISION-AWARE STRUCTURAL COMPRESSION

For the problem in Section 2.1, we represent uncertainty using a finite library $\mathcal{H}$. Each hypothesis h specifies a causal graph G_h, structural mechanisms, and a noise model, inducing an interventional distribution $P_h^x = P_h(\cdot \mid do(x))$ on the same rollout space as $P_x^\star$. We assume finite first moments of Y under the true system and every model. Each hypothesis specifies a causal graph, structural mechanisms, and a noise model; sequential observations update posterior weights over the resulting joint causal-model hypotheses.

At the beginning of round t, let $w_{t,h}$ be the normalized posterior weight of hypothesis h given the observed history $\mathcal{D}_t$. Since multiple mechanism hypotheses may share a graph, these weights induce a graph posterior and a graph-conditioned predictive distribution:

$$q_t(G) = \sum_{h:G_h=G} w_{t,h}, \qquad P_t(\cdot \mid x, G) = \sum_{h:G_h=G} \frac{w_{t,h}}{q_t(G)} P_h^x, \tag{2}$$

defined for $G \in \mathcal{G}_t^+ = \{G : q_t(G) > 0\}$. The decision maps are $J_t(x,G) = \mathbb{E}_{P_t(\cdot|x,G)}[Y]$ and $r_t(x,G) = \mathbb{E}_{P_t(\cdot|x,G)}[Z]$. Their q_t-weighted averages, denoted by $J_t(x)$ and $r_t(x)$, give the objective and risk under the full predictive mixture $P_t(\cdot \mid x) = \sum_h w_{t,h} P_h^x$.

At each round, the history determines a nonempty finite candidate set $\mathcal{U}_t \subseteq \mathcal{X}$ before simulation begins, and all subsequent action minimizations use finite enumeration. To estimate the graph-conditioned maps, we generate N_t^{pil} independent pilot rollouts for each $(u,G) \in \mathcal{U}_t \times \mathcal{G}_t^+$. Each

rollout resamples h with probability $w_{t,h}/q_t(G)$ among hypotheses with graph G, then simulates its SCM noise. The sample means of Y and Z yield $\widehat{J}_t(u,G)$ and $\widehat{r}_t(u,G)$.

We call two graphs decision-map equivalent at round t if their objective and risk maps agree on $\mathcal{U}_t$. To approximate this equivalence, we compare their pilot estimates through

$$\widehat{d}_t(G,G') = \max_{u\in\mathcal{U}_t}\left\{|\widehat{J}_t(u,G)-\widehat{J}_t(u,G')| + \zeta_{\text{comp}}|\widehat{r}_t(u,G)-\widehat{r}_t(u,G')|\right\}, \tag{3}$$

where $\zeta_{\text{comp}} > 0$ controls the relative contribution of risk differences. Starting from the graph with the largest posterior weight, farthest-first selection repeatedly adds the graph farthest from its nearest representative (Gonzalez, 1985). Selection stops when the covering radius is at most $\epsilon_t^{\text{cov}} \geq 0$ or the number of representatives reaches $m_t^{\max} \geq 1$. Ties are resolved by a fixed ordering. Let $\widehat{\mathcal{G}}_t$ be the resulting representative set and $T_t(G)$ the nearest representative of G. Assigning each graph's entire posterior mass to its representative gives

$$\begin{aligned} Q_t(g) &= \sum_{G:T_t(G)=g} q_t(G), \\ \widetilde{J}_t(u) &= \sum_g Q_t(g)\widehat{J}_t(u,g), \qquad \widetilde{r}_t(u) = \sum_g Q_t(g)\widehat{r}_t(u,g). \end{aligned} \tag{4}$$

The assignment preserves all posterior mass. Its effect on the decision maps is bounded as follows.

Proposition 3.1 (Decision-map approximation)**.** *For $v \in \{J, r\}$, let*

$$\rho_{t,v} = \max_{G,u}|\widehat{v}_t(u,G) - v_t(u,G)|, \qquad D_{t,v} = \sum_G q_t(G)\max_u|\widehat{v}_t(u,G) - \widehat{v}_t(u,T_t(G))|,$$

where $G \in \mathcal{G}_t^+$ and $u \in \mathcal{U}_t$. Then, for every candidate u,

$$|J_t(u) - \widetilde{J}_t(u)| \leq \rho_{t,J} + D_{t,J}, \qquad |r_t(u) - \widetilde{r}_t(u)| \leq \rho_{t,r} + D_{t,r}.$$

The bound holds for each realized pilot, including data-dependent representative selection, and separates estimation error $\rho_{t,v}$ from the computable distortion $D_{t,v}$. The compressed representation also reduces the cost of repeatedly evaluating downstream intervention decisions during SVI computation.

3.2 Structural Value of Information and Acquisition

The value of structural information depends on how it changes downstream intervention decisions, rather than simply on how much it reduces graph uncertainty. We define Structural Value of Information (SVI) as the expected reduction in optimal risk-penalized loss under a hypothetical graph-posterior update. Using the pilot estimates from Section 3.1, let

$$\ell_t(u,G) = \widehat{J}_t(u,G) + \zeta_{\text{info}}\widehat{r}_t(u,G), \qquad L_t^{\text{full}}(q) = \min_{u\in\mathcal{U}_t}\sum_G q(G)\ell_t(u,G). \tag{5}$$

Here $\zeta_{\text{info}} \geq 0$ weights predicted constraint violations in the hypothetical decision. During SVI evaluation, the loss maps remain fixed and mechanism uncertainty is marginalized within each graph-conditioned predictive law. The risk penalty shapes the information value; certification in Section 3.3 enforces execution feasibility. With the decision maps held fixed, L_t^{full} is concave in the graph posterior, which ensures that SVI is nonnegative under Bayesian reweighting.

Let $\mathcal{L}_h(o \mid do(x))$ be the likelihood of the recorded observation o under hypothesis h. The graph-conditioned and full predictive likelihoods are

$$\begin{aligned} p_t(o \mid x, G) &= \sum_{h:G_h=G}\frac{w_{t,h}}{q_t(G)}\mathcal{L}_h(o \mid do(x)), \\ p_t^{\text{full}}(o \mid x) &= \sum_G q_t(G)p_t(o \mid x, G). \end{aligned}$$

An observation with positive predictive likelihood induces $q_t^o(G) = q_t(G)p_t(o \mid x, G)/p_t^{\text{full}}(o \mid x)$. We define the full-library Structural Value of Information as the expected reduction in optimal loss after this update:

$$\text{SVI}_t^{\text{full}}(x) = L_t^{\text{full}}(q_t) - \mathbb{E}_{o\sim p_t^{\text{full}}(\cdot|x)}[L_t^{\text{full}}(q_t^o)].$$

To evaluate this value with fewer decision maps, we use the representatives from Section 3.1. Similar objective and risk maps, however, need not imply similar observation likelihoods. We therefore aggregate the likelihoods of all graphs within each cluster $\mathcal{C}_t(g) = \{G : T_t(G) = g\}$:

$$p_t^{\mathrm{cl}}(o \mid x, g) = \frac{1}{Q_t(g)} \sum_{G \in \mathcal{C}_t(g)} q_t(G) p_t(o \mid x, G),$$

$$Q_t^o(g) = \frac{Q_t(g) p_t^{\mathrm{cl}}(o \mid x, g)}{p_t^{\mathrm{full}}(o \mid x)}.$$

This construction preserves the full predictive law, $\sum_g Q_t(g) p_t^{\mathrm{cl}}(o \mid x, g) = p_t^{\mathrm{full}}(o \mid x)$, and gives $Q_t^o(g) = \sum_{G \in \mathcal{C}_t(g)} q_t^o(G)$. Using representative loss maps with these exact cluster weights yields

$$L_t^{\mathrm{cl}}(Q) = \min_{u \in \mathcal{U}_t} \sum_g Q(g) \ell_t(u, g),$$

$$\mathrm{SVI}_t^{\mathrm{cl}}(x) = L_t^{\mathrm{cl}}(Q_t) - \mathbb{E}_{o \sim p_t^{\mathrm{full}}(\cdot \mid x)}[L_t^{\mathrm{cl}}(Q_t^o)].$$

To relate this decision value to graph information and compression, let

$$I_t^{\mathrm{full}}(G; O_x) = \mathbb{E}_{o \sim p_t^{\mathrm{full}}(\cdot \mid x)} D_{\mathrm{KL}}(q_t^o \| q_t)$$

be the graph mutual information for the prospective observation O_x, using natural logarithms, and define the loss-map distortion

$$\Delta_t = \sum_G q_t(G) \max_{u \in \mathcal{U}_t} |\ell_t(u, G) - \ell_t(u, T_t(G))| \le D_{t,J} + \zeta_{\mathrm{info}} D_{t,r}.$$

Proposition 3.2 (Information value and compression error). *Let* $d_t^\ell = \max_{u \in \mathcal{U}_t}(\max_G \ell_t(u, G) - \min_G \ell_t(u, G))$, *where* $G \in \mathcal{G}_t^+$. *For every candidate* x,

$$0 \le \mathrm{SVI}_t^{\mathrm{full}}(x) \le d_t^\ell \min\left\{1, \sqrt{I_t^{\mathrm{full}}(G; O_x)/2}\right\},$$

$$|\mathrm{SVI}_t^{\mathrm{full}}(x) - \mathrm{SVI}_t^{\mathrm{cl}}(x)| \le 2\Delta_t.$$

The proof in Appendix C also gives $0 \le \mathrm{SVI}_t^{\mathrm{cl}}(x) \le d_t^\ell$. SVI vanishes when a common action remains optimal for almost every predictive observation, even if the observation is informative about the graph. Moreover, $\Delta_t = 0$ preserves full-library SVI exactly: graphs with identical loss maps can be merged without losing decision value, provided their predictive evidence is aggregated through the exact cluster likelihoods. Corollary B.2 further relates Δ_t to the covering tolerance, linking compression accuracy to representative selection.

For acquisition, we estimate the compressed SVI by Monte Carlo simulation. Conditional on the history and pilot, draw B_t independent observations $o_b \sim p_t^{\mathrm{full}}(\cdot \mid x)$ for each candidate x, evaluate each hypothetical optimal loss by finite enumeration, and set

$$\widehat{s}_t(x) = \left[L_t^{\mathrm{cl}}(Q_t) - B_t^{-1} \sum_{b=1}^{B_t} L_t^{\mathrm{cl}}(Q_t^{o_b}) \right]_{[0, d_t^\ell]},$$

where $[z]_{[a,b]} = \min\{b, \max\{a, z\}\}$. Appendix C bounds the Monte Carlo error conditional on the pilot. To account also for pilot error, let $\mathrm{SVI}_t^{\mathrm{ex}}$ use the exact predictive maps J_t, r_t in place of $\widehat{J}_t, \widehat{r}_t$, with the same graph update and fixed-map hypothetical decision. Under the pilot concentration condition, Corollary C.2 gives, with probability at least $1 - \delta_t^{\mathrm{pil}} - \delta_t^{\mathrm{svi}}$,

$$\max_{x \in \mathcal{U}_t} |\widehat{s}_t(x) - \mathrm{SVI}_t^{\mathrm{ex}}(x)| \le 2(a_{t,J} + \zeta_{\mathrm{info}} a_{t,r} + \Delta_t) + b_t, \tag{6}$$

where b_t is the Monte Carlo radius in equation C.1. This bound separates pilot estimation, structural compression, and Monte Carlo error within the finite predictive model. Combining this estimate with the compressed objective and a bounded intervention cost c gives the acquisition loss

$$\widehat{L}_t(x) = \widetilde{J}_t(x) - \lambda_t^{\mathrm{info}} \widehat{s}_t(x) + \gamma c(x), \tag{7}$$

with $\lambda_t^{\mathrm{info}}, \gamma \ge 0$. The score trades off predicted objective, structural information, and intervention cost; a larger SVI lowers the loss. We minimize it over the certified set defined next.

3.3 Safe Sequential Intervention and Posterior Updating

In sequential causal Bayesian optimization, interventions are executed repeatedly under evolving structural uncertainty, so safety must be enforced at each execution. For each candidate $u \in \mathcal{U}_t$, we draw N_t^{cert} fresh independent rollouts from the full posterior predictive mixture, resampling $h \sim w_t$ and its SCM noise on every draw. Let $\widehat{R}_t(u)$ denote the resulting sample mean of the joint violation indicator, whose posterior-predictive mean is $r_t(u)$. We then define the upper confidence bound

$$U_t^{\mathrm{risk}}(u) = \min\left\{1,\ \widehat{R}_t(u) + \eta_t^{\mathrm{cert}}\right\}. \tag{8}$$

Here $n_t = |\mathcal{U}_t|$ and $\delta_t^{\mathrm{cert}} \in (0,1)$ is the per-round confidence allocation. The radius $\eta_t^{\mathrm{cert}} = \sqrt{\log(n_t/\delta_t^{\mathrm{cert}})/(2N_t^{\mathrm{cert}})}$ follows from Hoeffding's inequality (Hoeffding, 1963) and a union bound over candidates. For a model-risk budget $0 \le \alpha_t \le \alpha$, the certified set and selection rule are

$$S_t = \{u \in \mathcal{U}_t : U_t^{\mathrm{risk}}(u) \le \alpha_t\}, \qquad x_t \in \arg\min_{u \in S_t} \widehat{L}_t(u). \tag{9}$$

The candidate set, budgets, and batch size are fixed before certification sampling. We use $\alpha_t = \alpha$ unless a model-discrepancy margin is available. Ties follow a fixed ordering, and the policy abstains when $S_t = \varnothing$.

For a selected action, we execute $do(x_t)$ and record o_t. Each intervention starts a separate episode from a common regime; conditional on the chosen action, its outcome follows $P^\star_{x_t}$ without carryover effects. The observation updates all hypothesis weights by

$$w_{t+1,h} = \frac{w_{t,h}\mathcal{L}_h(o_t \mid do(x_t))}{\sum_{b \in \mathcal{H}} w_{t,b}\mathcal{L}_b(o_t \mid do(x_t))}. \tag{10}$$

The history becomes $\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{(x_t, o_t)\}$, after which we rebuild the decision maps, representatives, and acquisition scores. Thus the actual update reweights both graph and mechanism hypotheses, whereas SVI evaluates structural reweighting with the current maps fixed. Appendix A specifies initialization and likelihood conditions.

Theorem 3.3 (Simultaneous posterior-predictive safety certification)**.** *Under the certification protocol above, for any horizon T such that $\sum_{t=1}^{T} \delta_t^{\mathrm{cert}} \le \delta_{\mathrm{cert}}$, with probability at least $1 - \delta_{\mathrm{cert}}$, every $u \in S_t$ satisfies $r_t(u) \le \alpha_t$, simultaneously for $t \le T$.*

The confidence level controls certification error, while α_t bounds the per-intervention violation probability. This risk averages over the current graph and mechanism posterior. Full-mixture certification makes this guarantee independent of compression error. Physical-risk control additionally depends on the accuracy of the predictive model.

Theorem 3.4 (Conditional physical-risk transfer)**.** *Under the conditions of Theorem 3.3, suppose that, on an event of probability at least $1 - \delta_{\mathrm{model}}$, $r^\star(u) \le r_t(u) + \epsilon_t^{\mathrm{phy}}$ for all candidates and rounds $t \le T$. Then, with probability at least $1 - \delta_{\mathrm{cert}} - \delta_{\mathrm{model}}$, all certified executed interventions satisfy $r^\star(x_t) \le \alpha_t + \epsilon_t^{\mathrm{phy}}$.*

The theorem shows that model-to-reality uncertainty can be accounted for by reserving a safety margin between the model-based risk threshold and the desired physical risk budget. Under additional model-accuracy and strict-feasibility conditions, Appendix E provides a conditional one-step objective-gap bound relative to the best sufficiently safe candidate, with a continuous-domain extension given in Appendix F.

4 Experiments

We evaluate DASCBO on a synthetic SCM and three application-derived SCMs, examining feasible optimization performance, empirical feasibility, and runtime. We compare against Causal Bayesian Optimization (CBO) (Aglietti et al., 2020), Constrained Causal Bayesian Optimization (CCBO) (Aglietti et al., 2023), Causal Bayesian Optimization with Unknown Graphs (CBOU) (Durand et al., 2025), Model-based Causal Bayesian Optimization (MCBO) (Sussex et al., 2023), and Causal Entropy Optimization (CEO) (Branchini et al., 2023). CBO, CCBO, and MCBO receive the reference

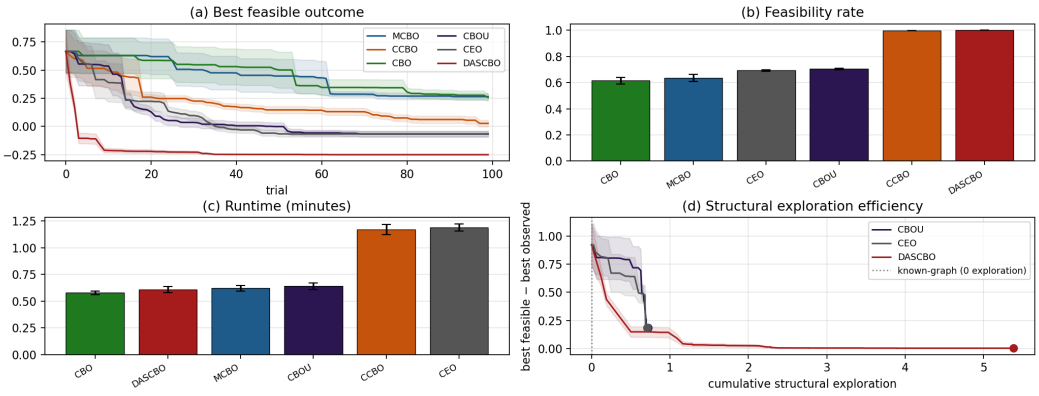


Figure 2: Results on the Synthetic Causal Optimization Benchmark. (a) Best feasible objective over intervention rounds. (b) Empirical feasibility rate. (c) Runtime per run in minutes on a linear scale. (d) Objective gap to the best observed feasible outcome versus cumulative target-parent posterior movement. Curves show means across 20 initializations; shaded regions indicate $\pm$ one standard error.

graph; CBOU, CEO, and DASCBO operate under structural uncertainty. All proposed interventions are assessed using the same benchmark feasibility criterion.

Protocol and metrics. All methods use the same initial observational and interventional sample sizes, with one initial point per intervention set. We use 20 initializations of $\mathcal{D}^I$, reporting the best feasible objective, the fraction of executed interventions satisfying all constraints, and runtime per run. Curves show means with $\pm$ one standard error across initializations.

Panel (d) plots the gap between the best feasible objective and the benchmark's best observed feasible outcome against cumulative target-parent posterior movement, measured by the sum of total-variation distances between successive parent-set posteriors. These trajectories relate feasible performance to changes in structural beliefs.

4.1 Synthetic Benchmark

The synthetic SCM minimizes Y subject to $X_4 \geq 0$, using intervention families $\mathcal{I} = \{\{X_1\}, \{X_2\}, \{X_3\}, \{X_1, X_2\}, \{X_1, X_3\}, \{X_2, X_3\}\}$. Figure 2(a) shows that DASCBO improves the feasible objective rapidly and attains the lowest final mean among the compared methods. DASCBO achieves 100% empirical feasibility, while CCBO has similarly high feasibility but settles at a higher objective. DASCBO's runtime is comparable to CBO, MCBO, and CBOU and lower than CCBO and CEO. Its posterior continues to change after CBOU and CEO plateau, accompanied by further objective improvements.

4.2 Application-Derived Causal Benchmarks

We next study healthcare, protein-signaling, and building-energy SCMs. These experiments evaluate simulated interventions in benchmark models informed by application data or prior models.

HEALTH. The HEALTH benchmark follows prior causal optimization work (Aglietti et al., 2023), with health-related relationships motivated by Ferro et al. (2015). It minimizes prostate-specific antigen (*PSA*) subject to *BMI* $<$ 25. Candidate intervention families are $\mathcal{I} = \{\{CI\}, \{Aspirin, CI\}, \{Statin, CI\}, \{Aspirin, Statin, CI\}\}$, where *CI* denotes the caloric-intake intervention. DASCBO improves the feasible objective rapidly and maintains an empirical feasibility rate of 100% (Figure 3). CCBO attains a slightly lower final objective and runs faster, but has a lower feasibility rate. DASCBO therefore achieves a comparable final objective with more consistent feasibility. CBOU has comparable runtime to DASCBO while returning a higher objective and substantially fewer feasible interventions.

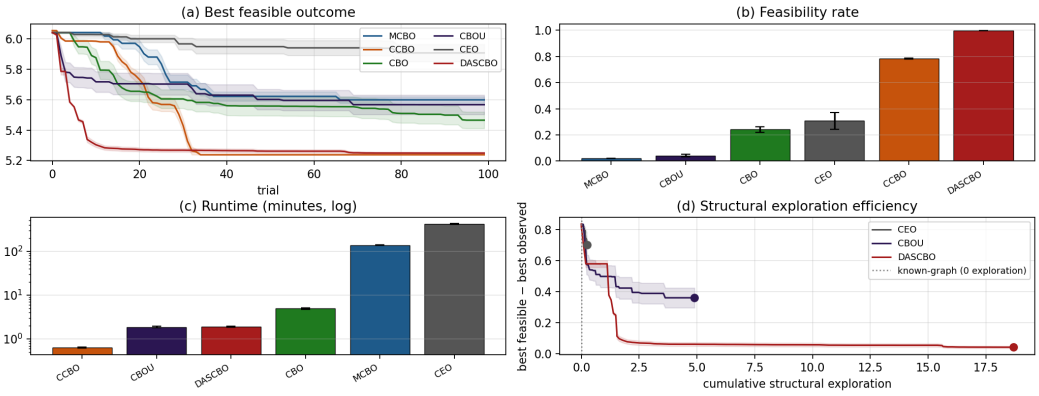


Figure 3: Results on the Healthcare Causal Optimization Benchmark. (a) Best feasible prostate-specific antigen objective over intervention rounds. (b) Empirical feasibility rate. (c) Runtime per run in minutes on a logarithmic scale. (d) Objective gap to the best observed feasible outcome versus cumulative target-parent posterior movement. Curves show means across 20 initializations; shaded regions indicate $\pm$ one standard error.

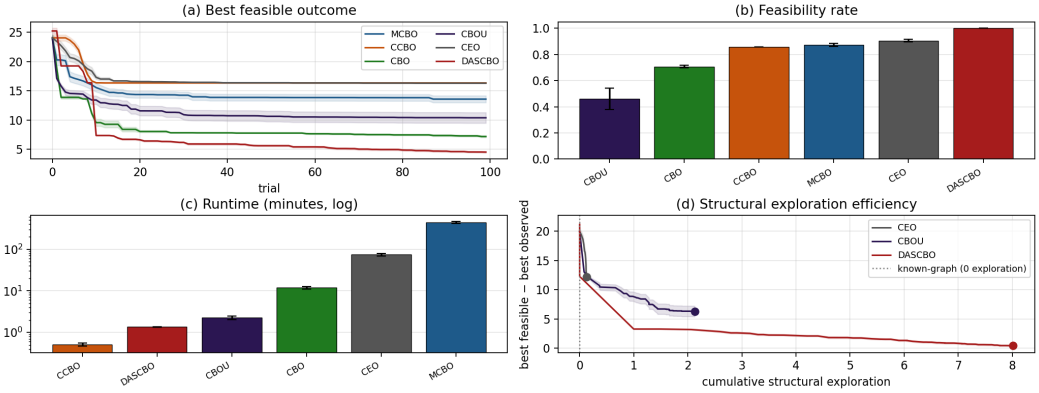


Figure 4: Results on the Protein-Signaling Causal Optimization Benchmark. (a) Best feasible *Erk* objective over intervention rounds. (b) Empirical feasibility rate. (c) Runtime per run in minutes on a logarithmic scale. (d) Objective gap to the best observed feasible outcome versus cumulative target-parent posterior movement. Curves show means across 20 initializations; shaded regions indicate $\pm$ one standard error.

PROTEIN-SIGNALING. This benchmark uses an SCM fitted to the protein-signaling data of Sachs et al. (2005), with *Erk* as the minimization target. Candidate intervention families are $\mathcal{I} = \{\{PKC\}, \{PKA\}, \{Mek\}, \{PKC, PKA\}, \{PKC, Mek\}, \{PKA, Mek\}\}$, with constraints on *PKC* and *PKA*. DASCBO makes a large early improvement and continues to reduce the feasible objective throughout the run (Figure 4). It achieves the lowest final mean objective and 100% empirical feasibility. CBO is the next-best method in final objective, but violates constraints more often. DASCBO also runs faster than the other methods that learn structure in this benchmark.

HONDA ENERGY SYSTEMS. The energy-management SCM is derived from the building dataset of Engel et al. (2025). It links environmental conditions, heating, cooling, combined-heat-and-power (CHP) operation, and electricity use. We minimize *grid electricity* using $\mathcal{I} = \{\{CHP\ heat\}, \{cooling\}, \{CHP\ heat, cooling\}\}$, subject to thresholds on *heating* and *cooling electricity*. DASCBO obtains the lowest final mean objective and an empirical feasibility rate of 0.994 (Figure 5). CCBO has high feasibility but makes little objective progress, whereas CBOU improves the objective with a substantially lower feasibility rate. DASCBO runs faster than CBO, MCBO, and CEO, although CCBO and CBOU remain faster.

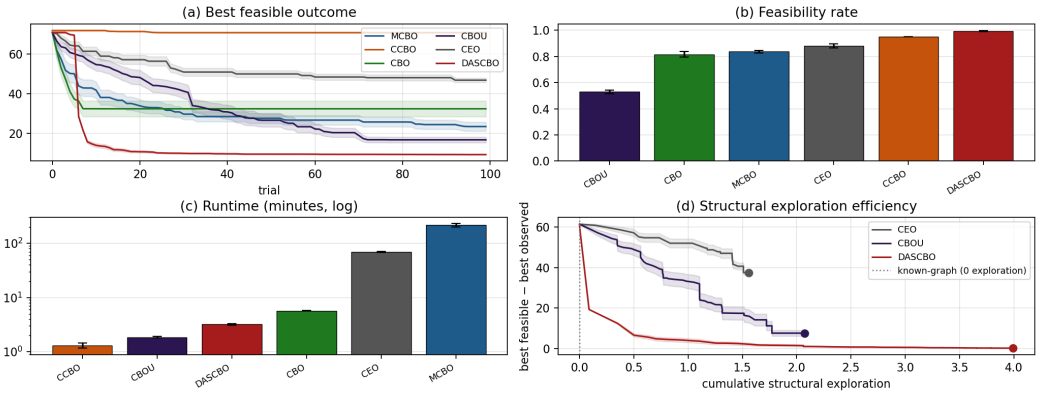


Figure 5: Results on the Honda Building Energy Management Benchmark. (a) Best feasible grid-electricity objective over intervention rounds. (b) Empirical feasibility rate. (c) Runtime per run in minutes on a logarithmic scale. (d) Objective gap to the best observed feasible outcome versus cumulative target-parent posterior movement. Curves show means across 20 initializations; shaded regions indicate $\pm$ one standard error.

These results show strong feasible-objective performance with high empirical feasibility across the four benchmarks. Appendix G analyzes the computational savings from decision-map compression.

Sensitivity analysis. On the synthetic benchmark, we vary structural-library size, constraint threshold, and risk tolerance with its certification budget. Across ten paired seeds, the tested settings reach similar final feasible objectives. Tighter constraints increase early variability, while lower risk tolerances require larger certification budgets. Appendix H reports the configurations and full results.

5 Conclusions

We introduced Decision-Aware Safe Causal Bayesian Optimization (DASCBO) for constrained intervention selection under structural uncertainty. Structural Value of Information (SVI) measures the expected improvement in the best risk-penalized decision after a graph-posterior update. This connects structural learning to the intervention objective: information is valuable when it changes the decisions supported by the current model.

DASCBO combines this criterion with decision-space compression and independent safety certification. Representative objective and risk maps reduce repeated decision computations, while full likelihood evidence preserves structural posterior updates. Our analysis establishes properties of SVI, bounds pilot, compression, and Monte Carlo errors, and gives simultaneous posterior-predictive safety guarantees. In particular, SVI vanishes when all graph hypotheses share an optimal action, even if an observation is informative about the graph. This property formalizes why structural information should be valued through its decision consequences. Model-discrepancy conditions further connect the safety guarantees to physical risk and conditional decision quality.

Experiments show strong feasible-objective performance and high empirical feasibility across four causal benchmarks. DASCBO achieves the lowest final mean objective on three benchmarks and a competitive objective with higher feasibility than CCBO on Healthcare. Its lower runtime than CEO on the application-derived benchmarks also demonstrates the practicality of combining decision-aware exploration with full-mixture risk certification in these settings. The sensitivity study finds similar final objectives across the tested library sizes and safety settings, with greater early variability under tighter constraints and larger certification budgets at lower risk tolerances. Together, these results support using downstream decision value to guide structural exploration while certifying execution risk separately. Extending the approach to larger model spaces, incorporating mechanism learning into the information criterion, and calibrating model discrepancy are promising directions for broader application.

AI USE STATEMENT

In this work, we used generative AI tools to aid and polish the writing of the paper and to support retrieval and discovery of related work, including identifying relevant literature. We have reviewed all AI-assisted work and take full responsibility for the final content of this submission, including all text, claims, and other materials.

REPRODUCIBILITY STATEMENT

Proofs of all theoretical results are provided in the Appendix. The implementation and source code accompanying this work are included in the Supplementary Material to facilitate reproduction of the reported results.

REFERENCES

Virginia Aglietti, Xiaoyu Lu, Andrei Paleyes, and Javier González. Causal Bayesian optimization. In *International Conference on Artificial Intelligence and Statistics*, volume 108, pp. 3155–3164. PMLR, 2020.

Virginia Aglietti, Alan Malek, Ira Ktena, and Silvia Chiappa. Constrained causal Bayesian optimization. In *International Conference on Machine Learning*, volume 202, pp. 304–321. PMLR, 2023.

Felix Berkenkamp, Andreas Krause, and Angela P. Schoellig. Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics. *Machine Learning*, 112:3713–3747, 2023. doi: 10.1007/s10994-021-06019-1.

Nicola Branchini, Virginia Aglietti, Neil Dhir, and Theodoros Damoulas. Causal entropy optimization. In *International Conference on Artificial Intelligence and Statistics*, volume 206, pp. 8586–8605. PMLR, 2023.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2nd edition, 2006. doi: 10.1002/047174882X.

Jean Durand, Yashas Annadani, Stefan Bauer, and Sonali Parbhoo. Causal Bayesian optimization with unknown graphs, 2025. arXiv:2503.19554.

Muhammad Qasim Elahi, Mahsa Ghasemi, and Murat Kocaoglu. Partial structure discovery is sufficient for no-regret learning in causal bandits. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Jens Engel, Andrea Castellani, Patricia Wollstadt, Felix Lanfermann, Thomas Schmitt, Sebastian Schmitt, Lydia Fischer, Steffen Limmer, David Luttropp, Florian Jomrich, René Unger, and Tobias Rodemann. A real-world energy management dataset from a smart company building for optimization and machine learning. *Scientific Data*, 12:864, 2025. doi: 10.1038/s41597-025-05186-3.

David Eriksson and Matthias Poloczek. Scalable constrained Bayesian optimization. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130, pp. 730–738. PMLR, 2021.

Ana Ferro, Francisco Pina, Milton Severo, Pedro Dias, Francisco Botelho, and Nuno Lunet. Use of statins and serum levels of prostate specific antigen. *Acta Urológica Portuguesa*, 32(2):71–77, 2015. doi: 10.1016/j.acup.2015.02.002.

Peter I. Frazier, Warren B. Powell, and Savas Dayanik. A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439, 2008. doi: 10.1137/070693424.

Peter I. Frazier, Warren B. Powell, and Savas Dayanik. The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21(4):599–613, 2009. doi: 10.1287/ijoc.1080.0314.

Jacob R. Gardner, Matt J. Kusner, Zhixiang Xu, Kilian Q. Weinberger, and John P. Cunningham. Bayesian optimization with inequality constraints. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32(2), pp. 937–945. PMLR, 2014.

Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985. doi: 10.1016/0304-3975(85)90224-5.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

Ronald A. Howard. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26, 1966. doi: 10.1109/TSSC.1966.300074.

Finnian Lattimore, Tor Lattimore, and Mark D. Reid. Causal bandits: Learning good interventions via causal inference. In *Advances in Neural Information Processing Systems*, volume 29, pp. 1189–1197, 2016.

Sanghack Lee and Elias Bareinboim. Structural causal bandits: Where to intervene? In *Advances in Neural Information Processing Systems*, volume 31, pp. 2573–2583, 2018.

Yangyi Lu, Amirhossein Meisami, Ambuj Tewari, and Zhenyu Yan. Regret analysis of bandit problems with causal background knowledge. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, volume 124, pp. 141–150. PMLR, 2020.

Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Causal bandits with unknown graph structure. In *Advances in Neural Information Processing Systems*, volume 34, pp. 24817–24828. Curran Associates, Inc., 2021.

Sumantrak Mukherjee, Mengyan Zhang, Seth Flaxman, and Sebastian Vollmer. Graph agnostic causal Bayesian optimisation. In *NeurIPS BDU Workshop*, 2024. Workshop paper.

Quoc Phong Nguyen, Wan Theng Ruth Chew, Le Song, Bryan Kian Hsiang Low, and Patrick Jaillet. Optimistic Bayesian optimization with unknown constraints. In *International Conference on Learning Representations*, 2024.

Min Woo Park, Andy Arditi, Elias Bareinboim, and Sanghack Lee. Structural causal bandits under Markov equivalence. In *Advances in Neural Information Processing Systems*, volume 38, pp. 98891–98939. Curran Associates, Inc., 2025. doi: 10.52202/085713-3304.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, 2017.

Karen Sachs, Omar Perez, Dana Pe'er, Douglas A. Lauffenburger, and Garry P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721): 523–529, 2005. doi: 10.1126/science.1105809.

Rajat Sen, Karthikeyan Shanmugam, Alexandros G. Dimakis, and Sanjay Shakkottai. Identifying best interventions through online importance sampling. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 3057–3066. PMLR, 2017.

Mohammad Shahverdikondori, Jalal Etesami, and Negar Kiyavash. Graph learning is suboptimal in causal bandits. In *International Conference on Artificial Intelligence and Statistics*, volume 300, pp. 3088–3096. PMLR, 2026.

Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*, volume 28 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 3rd edition, 2021. doi: 10.1137/1.9781611976595.

Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with Gaussian processes. In *International Conference on Machine Learning*, volume 37, pp. 997–1005. PMLR, 2015.

Scott Sussex, Anastasiia Makarova, and Andreas Krause. Model-based causal Bayesian optimization. In *International Conference on Learning Representations*, 2023.

Yunchuan Zhang, Sangwoo Park, and Osvaldo Simeone. Bayesian optimization with formal safety guarantees via online conformal prediction. *IEEE Journal of Selected Topics in Signal Processing*, 19(1):45–59, 2025. doi: 10.1109/JSTSP.2024.3422825.

A Sampling and Bayesian Update

We use the model library and notation of Section 3.1. The library and a strictly positive normalized prior are specified before the data used for posterior weighting. Let $\mathcal{D}_t$ contain any initial weighting data and the recorded intervention–observation pairs from rounds before t. For each action, model likelihoods are evaluable densities or mass functions with respect to a common dominating measure, obtained from the same laws P_h^x used for simulation. With fully observed SCMs and independent exogenous noises, the likelihood factors over non-intervened node conditionals; with partial observation, it is the marginal likelihood of the recorded variables.

Conditional on the observed history, the action-selection rule is the same under every library hypothesis, so its factor cancels in Bayes' rule, yielding equation 10. We compute normalized log weights for numerical stability. If an actual observation has zero predictive likelihood, the policy stops model-guided execution. An abstaining round leaves the observation history and posterior unchanged.

Each graph-conditioned rollout independently samples h with probability $w_{t,h}/q_t(G)$ among hypotheses with graph G, then samples that model's exogenous noise. A full-mixture rollout instead samples $h \sim w_t$. In both cases, the hypothesis and noise are resampled on every draw. Finite first moments ensure that the resulting finite pilot tables are almost surely finite.

The history determines $\mathcal{U}_t$ before pilot simulation. Conditional on the history and realized pilot, SVI samples are independent draws from the full predictive law. Let $\mathcal{F}_t^{\mathrm{pre}}$ contain the history, candidate set, pilot, SVI samples, and certification settings. Conditional on $\mathcal{F}_t^{\mathrm{pre}}$, each certification batch consists of i.i.d. draws from $P_t(\cdot \mid u)$, using fresh randomness. All batch sizes are positive integers fixed before their respective samples. Confidence allocations are deterministic, with $\delta_t^{\mathrm{pil}}, \delta_t^{\mathrm{svi}} \in (0,1)$; for the decision-quality bounds, take $\delta_t^{\mathrm{cert}} \in (0, 1/2)$. Ties in representative assignment and action selection follow a fixed ordering.

B Compression Error and Pilot Concentration

Proof of Proposition 3.1. By the mass assignment in equation 4, inserting the full weighted pilot average yields, for $v \in \{J, r\}$,

$$\begin{aligned} |v_t(u) - \widetilde{v}_t(u)| &\le \left| \sum_G q_t(G)\{v_t(u,G) - \widehat{v}_t(u,G)\} \right| \\ &\quad + \left| \sum_G q_t(G)\{\widehat{v}_t(u,G) - \widehat{v}_t(u, T_t(G))\} \right| \\ &\le \rho_{t,v} + D_{t,v} =: e_{t,v}. \end{aligned} \tag{B.1}$$

The inequality holds simultaneously for all candidates for each realized pilot and assignment T_t. □

To obtain a confidence bound for the pilot error, let $g_t = |\mathcal{G}_t^+|$, $n_t = |\mathcal{U}_t|$, and assume a known finite sub-Gaussian proxy $\bar{\sigma}_{t,J}$, fixed before pilot sampling, such that

$$\mathbb{E}_{P_t(\cdot|u,G)} e^{s(Y - J_t(u,G))} \le e^{s^2 \bar{\sigma}_{t,J}^2/2} \quad \text{for all } s \in \mathbb{R},\ u \in \mathcal{U}_t,\ G \in \mathcal{G}_t^+.$$

This condition is imposed on the graph-conditioned mixture, including variation between mechanism means. Define

$$a_{t,J} = \bar{\sigma}_{t,J} \sqrt{\frac{2\log(4 g_t n_t / \delta_t^{\mathrm{pil}})}{N_t^{\mathrm{pil}}}}, \qquad a_{t,r} = \sqrt{\frac{\log(4 g_t n_t / \delta_t^{\mathrm{pil}})}{2 N_t^{\mathrm{pil}}}}. \tag{B.2}$$

Corollary B.1 (Pilot confidence bound)**.** *Under the sub-Gaussian condition and the sampling protocol in Appendix A, the event*

$$E_t^{\mathrm{pil}} = \{\rho_{t,J} \le a_{t,J},\ \rho_{t,r} \le a_{t,r}\}$$

has probability at least $1 - \delta_t^{\mathrm{pil}}$ *conditional on the history. On this event,*

$$e_{t,J} \le \bar{e}_{t,J} := a_{t,J} + D_{t,J}, \qquad e_{t,r} \le \bar{e}_{t,r} := a_{t,r} + D_{t,r}.$$

Proof. For $\bar{\sigma}_{t,J} > 0$, the two-sided objective and risk tails are bounded by $2\exp[-N_t^{\mathrm{pil}} a_{t,J}^2/(2\bar{\sigma}_{t,J}^2)]$ and $2\exp[-2N_t^{\mathrm{pil}} a_{t,r}^2]$, respectively. Each equals $\delta_t^{\mathrm{pil}}/(2g_t n_t)$. A union bound over both quantities and all graph–candidate pairs proves the claim. If $\bar{\sigma}_{t,J} = 0$, the objective estimation error is zero almost surely. □

The same bounds relate empirical coverage to coverage of the exact maps. Let d_t denote equation 3 evaluated with J_t, r_t. Then

$$d_t(G, G') \le \widehat{d}_t(G, G') + 2\rho_{t,J} + 2\zeta_{\mathrm{comp}}\rho_{t,r}.$$

When selection reaches the coverage tolerance, the exact-map radius is at most $\epsilon_t^{\mathrm{cov}} + 2a_{t,J} + 2\zeta_{\mathrm{comp}} a_{t,r}$ on E_t^{pil}. If selection stops at the representative cap, the realized distortions $D_{t,J}, D_{t,r}$ still give the bounds in Proposition 3.1.

For a risk budget b, define $V_t(b) = \min_{u: r_t(u) \le b} J_t(u)$ and $\widetilde{V}_t(b) = \min_{u: \widetilde{r}_t(u) \le b} \widetilde{J}_t(u)$, with $u \in \mathcal{U}_t$. If the invoked feasible sets are nonempty and $b - e_{t,r}, b, b + e_{t,r} \in [0, 1]$, the map bounds imply

$$\widetilde{V}_t(b + e_{t,r}) - e_{t,J} \le V_t(b) \le \widetilde{V}_t(b - e_{t,r}) + e_{t,J}. \tag{B.3}$$

Indeed, exact feasibility at b implies compressed feasibility at $b + e_{t,r}$, while compressed feasibility at $b - e_{t,r}$ implies exact feasibility at b. Applying the objective error bound gives the two inequalities.

Corollary B.2 (Coverage control of SVI distortion)**.** *Let $R_t^{\mathrm{cov}} = \max_{G \in \mathcal{G}_t^+} \widehat{d}_t(G, T_t(G))$ be the realized covering radius and $K_\zeta = \max\{1, \zeta_{\mathrm{info}}/\zeta_{\mathrm{comp}}\}$. Then*

$$\Delta_t \le K_\zeta \sum_G q_t(G) \widehat{d}_t(G, T_t(G)) \le K_\zeta R_t^{\mathrm{cov}}. \tag{B.4}$$

If the coverage criterion is met, then $\Delta_t \le K_\zeta \epsilon_t^{\mathrm{cov}}$, so Proposition 3.2 bounds the SVI compression error by $2K_\zeta \epsilon_t^{\mathrm{cov}}$. If the representative cap is reached first, the bound using R_t^{cov} still applies.

Proof. For each graph–candidate pair, the triangle inequality gives

$$\begin{aligned} |\ell_t(u, G) - \ell_t(u, T_t(G))| &\le |\widehat{J}_t(u, G) - \widehat{J}_t(u, T_t(G))| \\ &\quad + \zeta_{\mathrm{info}} |\widehat{r}_t(u, G) - \widehat{r}_t(u, T_t(G))| \\ &\le K_\zeta \widehat{d}_t(G, T_t(G)). \end{aligned}$$

Taking the maximum over candidates and averaging under q_t proves the claim. □

Appendix G summarizes the pilot, clustering, and SVI evaluation costs.

C SVI Properties and Monte Carlo Error

Expectations over prospective observations use $p_t^{\mathrm{full}}(\cdot \mid x)$. On its null set, define $q_t^o = q_t$ and $Q_t^o = Q_t$.

Proof of Proposition 3.2. The functions L_t^{full} and L_t^{cl} are finite minima of linear functions and hence concave. Bayes' rule gives $\mathbb{E}_o q_t^o = q_t$ and $\mathbb{E}_o Q_t^o = Q_t$, so Jensen's inequality implies nonnegative SVI in both cases. For any graph distributions q, q',

$$\begin{aligned} |L_t^{\mathrm{full}}(q) - L_t^{\mathrm{full}}(q')| &\le \max_u \left| \sum_G (q(G) - q'(G)) \ell_t(u, G) \right| \\ &\le d_t^\ell d_{\mathrm{TV}}(q, q'), \qquad d_{\mathrm{TV}}(q, q') = \tfrac{1}{2} \sum_G |q(G) - q'(G)|. \end{aligned}$$

The second inequality follows from the range of each loss map across graphs. Pinsker's inequality (Cover & Thomas, 2006) and Jensen's inequality then give

$$\mathrm{SVI}_t^{\mathrm{full}}(x) \le d_t^\ell \sqrt{\tfrac{1}{2} \mathbb{E}_o D_{\mathrm{KL}}(q_t^o \| q_t)} = d_t^\ell \sqrt{\tfrac{1}{2} I_t^{\mathrm{full}}(G; O_x)}.$$

Since $d_{\mathrm{TV}} \le 1$, we also have $\mathrm{SVI}_t^{\mathrm{full}}(x) \le d_t^\ell$. The same argument for the representative loss table gives $0 \le \mathrm{SVI}_t^{\mathrm{cl}}(x) \le d_t^\ell$.

For the compression error, write $a_t(G) = \max_u |\ell_t(u, G) - \ell_t(u, T_t(G))|$. Given a graph distribution q and its cluster aggregation $Q(g) = \sum_{G \in \mathcal{C}_t(g)} q(G)$, the minimum inequality yields

$$|L_t^{\mathrm{full}}(q) - L_t^{\mathrm{cl}}(Q)| \le \sum_G q(G) a_t(G).$$

Exact cluster likelihoods ensure that Q_t^o is the aggregation of q_t^o. Applying this bound before and after the observation gives

$$\begin{aligned} |\mathrm{SVI}_t^{\mathrm{full}}(x) - \mathrm{SVI}_t^{\mathrm{cl}}(x)| &\le \Delta_t + \mathbb{E}_o \sum_G q_t^o(G) a_t(G) \\ &= 2\Delta_t, \end{aligned}$$

where the equality uses $\mathbb{E}_o q_t^o(G) = q_t(G)$. Finally, the triangle inequality gives $\Delta_t \le D_{t,J} + \zeta_{\mathrm{info}} D_{t,r}$. □

Corollary C.1 (Exact preservation under loss-map equivalence)**.** *If every graph and its representative have identical penalized loss maps on $\mathcal{U}_t$, then $\Delta_t = 0$ and*

$$\mathrm{SVI}_t^{\mathrm{full}}(x) = \mathrm{SVI}_t^{\mathrm{cl}}(x) \qquad \textit{for every } x \in \mathcal{U}_t.$$

Thus exact cluster likelihoods preserve SVI under loss-map equivalence even when the merged graphs have different observation likelihoods.

To verify the zero-value property, suppose u_0 minimizes the compressed posterior loss for almost every observation. Then

$$\mathbb{E}_o L_t^{\mathrm{cl}}(Q_t^o) = \sum_g Q_t(g) \ell_t(u_0, g) \ge L_t^{\mathrm{cl}}(Q_t) \ge \mathbb{E}_o L_t^{\mathrm{cl}}(Q_t^o),$$

where the last inequality is Jensen's. Equality follows, so $\mathrm{SVI}_t^{\mathrm{cl}}(x) = 0$. The same argument applies to full-library SVI.

For Monte Carlo estimation, the Lipschitz bound above implies $\sup_Q L_t^{\mathrm{cl}}(Q) - \inf_Q L_t^{\mathrm{cl}}(Q) \le d_t^\ell$. Consequently, the per-observation loss reductions have support width at most d_t^ℓ, although individual reductions can be negative. Hoeffding's inequality and a union bound over the $n_t = |\mathcal{U}_t|$ candidates give

$$\begin{aligned} E_t^{\mathrm{svi}} &= \left\{ \max_{x \in \mathcal{U}_t} |\widehat{s}_t(x) - \mathrm{SVI}_t^{\mathrm{cl}}(x)| \le b_t \right\}, \\ b_t &= d_t^\ell \sqrt{\frac{\log(2n_t/\delta_t^{\mathrm{svi}})}{2B_t}}, \end{aligned} \tag{C.1}$$

with conditional probability at least $1 - \delta_t^{\mathrm{svi}}$. Clipping to $[0, d_t^\ell]$ cannot increase the estimation error. If $d_t^\ell = 0$, both SVI and its estimate vanish. The bound is conditional on the fixed empirical loss table.

To include pilot estimation error, define the exact predictive loss and its structural value by

$$\begin{aligned} \ell_t^{\mathrm{ex}}(u, G) &= J_t(u, G) + \zeta_{\mathrm{info}} r_t(u, G), \\ L_t^{\mathrm{ex}}(q) &= \min_{u \in \mathcal{U}_t} \sum_G q(G) \ell_t^{\mathrm{ex}}(u, G), \\ \mathrm{SVI}_t^{\mathrm{ex}}(x) &= L_t^{\mathrm{ex}}(q_t) - \mathbb{E}_o[L_t^{\mathrm{ex}}(q_t^o)]. \end{aligned}$$

The maps remain fixed during the hypothetical update, and the expectation uses the same full predictive law as $\mathrm{SVI}_t^{\mathrm{full}}$. Thus $\mathrm{SVI}_t^{\mathrm{ex}}$ is the exact-map reference within the finite library.

Corollary C.2 (Total SVI approximation error)**.** *For every realized pilot and candidate x,*

$$\begin{aligned} |\mathrm{SVI}_t^{\mathrm{ex}}(x) - \mathrm{SVI}_t^{\mathrm{full}}(x)| &\le 2(\rho_{t,J} + \zeta_{\mathrm{info}} \rho_{t,r}), \\ |\mathrm{SVI}_t^{\mathrm{ex}}(x) - \mathrm{SVI}_t^{\mathrm{cl}}(x)| &\le 2(\rho_{t,J} + \zeta_{\mathrm{info}} \rho_{t,r} + \Delta_t). \end{aligned} \tag{C.2}$$

On $E_t^{\mathrm{pil}} \cap E_t^{\mathrm{svi}}$, the estimator satisfies equation 6 simultaneously for all candidates. Under the condition of Corollary B.1, this event has probability at least $1 - \delta_t^{\mathrm{pil}} - \delta_t^{\mathrm{svi}}$ conditional on the history.

Proof. For any graph distribution q, the minimum inequality implies

$$|L_t^{\mathrm{ex}}(q) - L_t^{\mathrm{full}}(q)| \le \max_{u,G} |\ell_t^{\mathrm{ex}}(u, G) - \ell_t(u, G)| \le \rho_{t,J} + \zeta_{\mathrm{info}} \rho_{t,r}.$$

Applying this inequality at q_t and at each q_t^o, then taking the observation expectation, proves the first bound. Proposition 3.2 gives the second. Combining it with equation C.1 and the pilot confidence radii gives equation 6. The Monte Carlo guarantee holds conditional on every realized pilot; integrating over the pilot and taking a union bound yields the stated probability. □

Together with equation B.4, this result relates the SVI error to the pilot and Monte Carlo sample sizes and the realized covering radius. When the coverage criterion is met, the compression term can be replaced by $K_\zeta \epsilon_t^{\mathrm{cov}}$.

D Safety Certification

Proof of Theorem 3.3. Conditional on $\mathcal{F}_t^{\mathrm{pre}}$, the certification indicators for candidate u are i.i.d. Bernoulli with mean $r_t(u)$. A one-sided Hoeffding bound and a union bound over candidates yield

$$\Pr\{\exists u \in \mathcal{U}_t : r_t(u) > U_t^{\mathrm{risk}}(u) \mid \mathcal{F}_t^{\mathrm{pre}}\} \le \delta_t^{\mathrm{cert}}.$$

Clipping the upper bound at one preserves validity because $r_t(u) \le 1$. Integrating over $\mathcal{F}_t^{\mathrm{pre}}$ and summing the failure probabilities over rounds proves simultaneous coverage, including for adaptively selected candidates. □

For example, the allocation $\delta_t^{\mathrm{cert}} = 6\delta_{\mathrm{cert}}/(\pi^2 t^2)$ satisfies $\sum_{t=1}^{\infty} \delta_t^{\mathrm{cert}} = \delta_{\mathrm{cert}}$, giving simultaneous certification for all rounds. For a desired radius $\eta > 0$, taking $N_t^{\mathrm{cert}} \ge \log(n_t/\delta_t^{\mathrm{cert}})/(2\eta^2)$ ensures $\eta_t^{\mathrm{cert}} \le \eta$. The union bound allows dependence across candidate batches.

Proof of Theorem 3.4. On the intersection of the simultaneous certification and model-accuracy events,

$$r^\star(x_t) \le r_t(x_t) + \epsilon_t^{\mathrm{phy}} \le \alpha_t + \epsilon_t^{\mathrm{phy}}$$

for every certified executed intervention. A union bound gives intersection probability at least $1 - \delta_{\mathrm{cert}} - \delta_{\mathrm{model}}$. □

A sufficient condition for transfer is $d_{\mathrm{TV}}(P_u^\star, P_t(\cdot \mid u)) \le \epsilon_t^{\mathrm{phy}}$ uniformly over candidates and rounds, since risk is the probability of a measurable event.

For the decision-quality analysis, we also need a two-sided certification event. With $0 < \delta_t^{\mathrm{cert}} < 1/2$,

$$E_t^{\mathrm{acc}} = \left\{ \max_{u \in \mathcal{U}_t} |\widehat{R}_t(u) - r_t(u)| \le \eta_t^{\mathrm{cert}} \right\}, \qquad \Pr(E_t^{\mathrm{acc}} \mid \mathcal{F}_t^{\mathrm{pre}}) \ge 1 - 2\delta_t^{\mathrm{cert}}. \tag{D.1}$$

On this event, $r_t(u) \le U_t^{\mathrm{risk}}(u) \le r_t(u) + 2\eta_t^{\mathrm{cert}}$. Hence any candidate satisfying $r_t(u) + 2\eta_t^{\mathrm{cert}} \le \alpha_t$ belongs to S_t, and every selected candidate has model risk at most α_t.

E Conditional Decision Quality

Let $e_{t,J} = \rho_{t,J} + D_{t,J}$ as in equation B.1. When $S_t \neq \varnothing$, the selection rule in equation 9 gives

$$\widehat{L}_t(x_t) \le \widehat{L}_t(u) \qquad \text{for all } u \in S_t. \tag{E.1}$$

Define the model-accuracy event

$$E_t^{\mathrm{dec}} = \left\{ \begin{array}{l} \max_{u \in \mathcal{U}_t} |J_t(u) - J^\star(u)| \le \epsilon_t^J, \\ \max_{u \in \mathcal{U}_t} |r_t(u) - r^\star(u)| \le \epsilon_t^r. \end{array} \right\}. \tag{E.2}$$

To define a meaningful benchmark, let

$$\mathcal{U}_t^{\mathrm{cmp}} = \left\{ u \in \mathcal{U}_t : r^\star(u) + \epsilon_t^r + 2\eta_t^{\mathrm{cert}} \le \alpha_t \right\} \tag{E.3}$$

be the set of candidates with sufficient safety margin. Whenever $\mathcal{U}_t^{\mathrm{cmp}} \neq \varnothing$, define

$$u_t^\dagger \in \arg\min_{u \in \mathcal{U}_t^{\mathrm{cmp}}} J^\star(u). \tag{E.4}$$

The comparator $u_t^\dagger$ is used only for analysis and need not be known to the learner.

Theorem E.1 (Conditional decision-gap bound)**.** *On $E_t^{\mathrm{acc}} \cap E_t^{\mathrm{dec}}$, suppose that $\mathcal{U}_t^{\mathrm{cmp}} \neq \varnothing$. Then $u_t^\dagger \in S_t$, and the selected intervention satisfies*

$$r^\star(x_t) \le \alpha_t + \epsilon_t^r.$$

Moreover,

$$J^\star(x_t) - J^\star(u_t^\dagger) \le 2(\epsilon_t^J + e_{t,J}) + \lambda_t^{\mathrm{info}} d_t^\ell + \gamma \operatorname{osc}_{\mathcal{U}_t} c, \tag{11}$$

where

$$\operatorname{osc}_{\mathcal{U}_t} c = \max_{u \in \mathcal{U}_t} c(u) - \min_{u \in \mathcal{U}_t} c(u).$$

Proof. Since $u_t^\dagger \in \mathcal{U}_t^{\mathrm{cmp}}$, on E_t^{dec},

$$r_t(u_t^\dagger) \le r^\star(u_t^\dagger) + \epsilon_t^r \le \alpha_t - 2\eta_t^{\mathrm{cert}}.$$

On E_t^{acc}, we have

$$U_t^{\mathrm{risk}}(u_t^\dagger) \le r_t(u_t^\dagger) + 2\eta_t^{\mathrm{cert}} \le \alpha_t,$$

so $u_t^\dagger \in S_t$. Hence $S_t \neq \varnothing$.

For the selected intervention $x_t \in S_t$, the event E_t^{acc} gives

$$r_t(x_t) \le U_t^{\mathrm{risk}}(x_t) \le \alpha_t.$$

Combining this with E_t^{dec} yields

$$r^\star(x_t) \le r_t(x_t) + \epsilon_t^r \le \alpha_t + \epsilon_t^r.$$

Since $u_t^\dagger \in S_t$, the selection rule equation E.1 implies

$$\widehat{L}_t(x_t) \le \widehat{L}_t(u_t^\dagger).$$

Substituting the acquisition loss equation 7 gives

$$\begin{aligned} \widetilde{J}_t(x_t) - \widetilde{J}_t(u_t^\dagger) &\le \lambda_t^{\mathrm{info}} \left\{ \widehat{s}_t(x_t) - \widehat{s}_t(u_t^\dagger) \right\} + \gamma \left\{ c(u_t^\dagger) - c(x_t) \right\} \\ &\le \lambda_t^{\mathrm{info}} d_t^\ell + \gamma \operatorname{osc}_{\mathcal{U}_t} c, \end{aligned}$$

where the second inequality uses $0 \le \widehat{s}_t(u) \le d_t^\ell$.

Finally, on E_t^{dec}, Proposition 3.1 gives

$$|J^\star(u) - \widetilde{J}_t(u)| \le |J^\star(u) - J_t(u)| + |J_t(u) - \widetilde{J}_t(u)| \le \epsilon_t^J + e_{t,J}$$

for every $u \in \mathcal{U}_t$. Applying this bound at both x_t and $u_t^\dagger$, and combining it with the preceding inequality, proves equation 11. □

Integrating the conditional guarantee in equation D.1 gives

$$\Pr(E_t^{\mathrm{acc}}) \geq 1 - 2\delta_t^{\mathrm{cert}}.$$

If

$$\Pr(E_t^{\mathrm{dec}}) \geq 1 - \delta_t^{\mathrm{dec}},$$

then Theorem E.1 holds with probability at least

$$1 - 2\delta_t^{\mathrm{cert}} - \delta_t^{\mathrm{dec}},$$

whenever $\mathcal{U}_t^{\mathrm{cmp}} \neq \varnothing$.

Under Corollary B.1, $e_{t,J}$ can be replaced by the computable bound

$$\bar{e}_{t,J} = a_{t,J} + D_{t,J},$$

with validity probability at least

$$1 - \delta_t^{\mathrm{pil}} - 2\delta_t^{\mathrm{cert}} - \delta_t^{\mathrm{dec}}.$$

For $\alpha_t = \alpha$, any candidate satisfying

$$r^\star(u) \leq \alpha - \kappa_t$$

belongs to $\mathcal{U}_t^{\mathrm{cmp}}$ whenever

$$\kappa_t \geq \epsilon_t^r + 2\eta_t^{\mathrm{cert}}.$$

If $\alpha_t = \alpha - \bar{\epsilon}_t^{\mathrm{phy}}$, the required safety margin increases by $\bar{\epsilon}_t^{\mathrm{phy}}$.

The decision-gap bound separates model and compression errors from the immediate effects of structural exploration and intervention cost. The Monte Carlo analysis in Appendix C separately controls how accurately the information bonus estimates SVI.

F Continuous-Domain Comparison

For a compact intervention domain $\mathcal{X} \subset \mathbb{R}^d$, assume that $J^\star$ and $r^\star$ are Lipschitz continuous with constants $L_J^\star$ and $L_r^\star$, respectively. Define the candidate fill distance and a strict-margin optimum by

$$h_t = \sup_{x \in \mathcal{X}} \min_{u \in \mathcal{U}_t} \|x - u\|, \qquad x_\Delta^\star \in \arg\min_{x \in \mathcal{X}:\, r^\star(x) \leq \alpha - \Delta} J^\star(x),$$

where $\Delta > 0$ and the strict-margin feasible set is nonempty.

Corollary F.1 (Continuous-domain comparison under a strict feasibility margin)**.** *On $E_t^{\mathrm{acc}} \cap E_t^{\mathrm{dec}}$, if*

$$\alpha - \Delta + L_r^\star h_t + \epsilon_t^r + 2\eta_t^{\mathrm{cert}} \leq \alpha_t,$$

then

$$\begin{aligned} J^\star(x_t) - J^\star(x_\Delta^\star) \leq 2(\epsilon_t^J + e_{t,J}) + \lambda_t^{\mathrm{info}} d_t^\ell \\ + \gamma\, \mathrm{osc}_{\mathcal{U}_t} c + L_J^\star h_t. \end{aligned}$$

Proof. Compactness of $\mathcal{X}$, continuity of $J^\star$ and $r^\star$, and nonemptiness of the strict-margin feasible set ensure the existence of $x_\Delta^\star$.

Let $\bar{u}_t \in \mathcal{U}_t$ be a nearest candidate to $x_\Delta^\star$. By the definition of h_t,

$$\|\bar{u}_t - x_\Delta^\star\| \leq h_t.$$

Lipschitz continuity therefore gives

$$r^\star(\bar{u}_t) \leq r^\star(x_\Delta^\star) + L_r^\star h_t \leq \alpha - \Delta + L_r^\star h_t,$$

and

$$J^\star(\bar{u}_t) \leq J^\star(x_\Delta^\star) + L_J^\star h_t.$$

The assumed margin implies

$$r^\star(\bar{u}_t) + \epsilon_t^r + 2\eta_t^{\mathrm{cert}} \leq \alpha_t,$$

so $\bar{u}_t \in \mathcal{U}_t^{\mathrm{cmp}}$ and the comparator set is nonempty. By the definition of $u_t^\dagger$,

$$J^\star(u_t^\dagger) \le J^\star(\bar{u}_t) \le J^\star(x_\Delta^\star) + L_J^\star h_t.$$

Applying Theorem E.1 yields

$$J^\star(x_t) - J^\star(u_t^\dagger) \le 2(\epsilon_t^J + e_{t,J}) + \lambda_t^{\mathrm{info}} d_t^\ell + \gamma \operatorname{osc}_{\mathcal{U}_t} c.$$

Combining the last two inequalities proves the result. □

This comparison is relative to the optimum under the stricter risk budget $\alpha - \Delta$. Comparing directly with the optimum under budget α additionally requires controlling the objective loss induced by tightening the risk constraint.

G Computational Complexity Analysis

Let $n_t = |\mathcal{U}_t|$, $g_t = |\mathcal{G}_t^+|$, $m_t = |\widehat{\mathcal{G}}_t|$, and $H_t = |\{h : w_{t,h} > 0\}|$. We compare compressed SVI with full-library SVI using the same pilot tables and sampling budgets. Table 1 reports per-round costs across all n_t acquisition candidates. Rollout and likelihood counts are reported separately because their costs depend on the SCM and observation dimension.

Stage and counting unit	Full-library SVI	Compressed SVI
Pilot rollouts	$n_t g_t N_t^{\mathrm{pil}}$	$n_t g_t N_t^{\mathrm{pil}}$
Clustering arithmetic	—	$O(g_t^2 n_t + g_t m_t)$
SVI predictive rollouts	$n_t B_t$	$n_t B_t$
SVI model likelihood evaluations	$n_t B_t H_t$	$n_t B_t H_t$
Hypothetical decision arithmetic	$O(B_t n_t^2 g_t)$	$O(B_t n_t^2 m_t)$
Certification rollouts	$n_t N_t^{\mathrm{cert}}$	$n_t N_t^{\mathrm{cert}}$
Posterior-update likelihood evaluations	H_t	H_t

Table 1: Per-round computational costs. The full-library baseline omits clustering; both methods retain the full predictive law and use the same certification and posterior-update procedures.

Computing all pairwise decision-map distances costs $O(g_t^2 n_t)$. Given the resulting distance matrix, farthest-first selection can maintain, for each graph, its distance to the nearest representative. Each newly selected representative then requires $O(g_t)$ updates, giving $O(g_t m_t)$ traversal cost.

For each prospective observation, both methods evaluate the likelihood under all H_t positive-weight hypotheses. Aggregating these contributions into graph or cluster weights is linear in H_t with stored membership indices. Compression therefore retains the full evidence computation.

For one acquisition candidate and one hypothetical observation, minimizing the posterior loss over n_t actions costs $O(n_t g_t)$ with the full graph table and $O(n_t m_t)$ with representative maps. Repeating this calculation over B_t observations and n_t acquisition candidates gives the hypothetical decision costs in Table 1.

Compression therefore reduces the repeated hypothetical decision arithmetic from $O(B_t n_t^2 g_t)$ to $O(B_t n_t^2 m_t)$, while adding the one-time clustering cost. The reduction in this component is by a factor of g_t/m_t; the effect on total runtime depends on the relative costs of simulation, likelihood evaluation, clustering, and decision enumeration. Corollary B.2 relates the SVI approximation error to the realized covering radius, quantifying the accuracy–compression tradeoff induced by representative selection.

H Sensitivity Analysis on the Synthetic Causal Optimization Benchmark

We vary the structural hypothesis library size, the constraint threshold, and the safety operating point. Each configuration uses the same ten paired seeds, with the initial intervention realization matched by seed index. All settings outside the dimension being varied remain fixed at their defaults. The safety operating-point study varies the risk tolerance together with its certification budget.

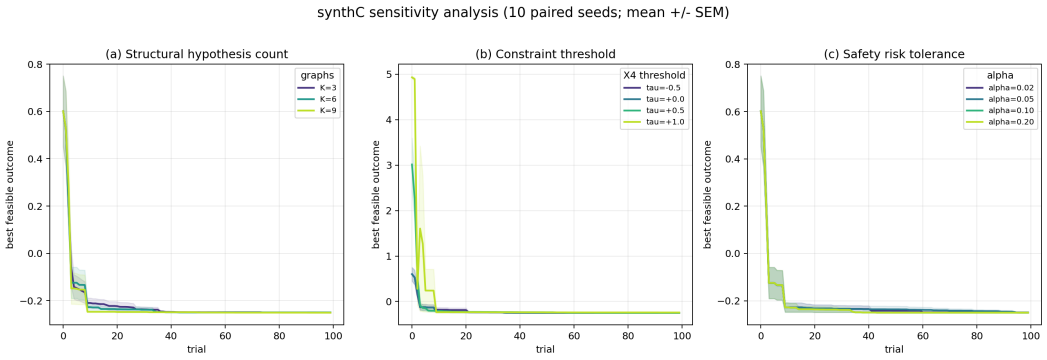


Figure 6: Sensitivity analysis on the Synthetic Causal Optimization Benchmark. Best feasible objective over intervention rounds for (a) structural hypothesis counts $K \in \{3, 6, 9\}$; (b) constraint thresholds $\tau \in \{-0.5, 0, 0.5, 1.0\}$ in $X_4 \geq \tau$; and (c) risk tolerances $\alpha \in \{0.02, 0.05, 0.10, 0.20\}$ with corresponding certification budgets. Curves show means across ten paired seeds; shaded regions indicate $\pm$ one standard error. Lower objectives are better.

Structural hypothesis library size. The three library sizes produce modest differences in early optimization and converge to similar final feasible objectives (Figure 6(a)). The mean empirical feasibility rate is 1.0 for each setting. Thus, comparable final performance is obtained throughout the tested range, including with the smallest library of three hypotheses.

Constraint threshold. Increasing τ from -0.5 to 1.0 tightens the constraint $X_4 \geq \tau$ and changes the feasible region. The strictest threshold yields higher and more variable early objectives, followed by a rapid reduction (Figure 6(b)). All four settings reach similar final feasible objectives under their respective constraints. The main observed effect of tightening the threshold is therefore on the early search trajectory.

Safety operating point. The four risk tolerances and their associated certification budgets lead to similar objective trajectories and a mean empirical feasibility rate of 1.0 (Figure 6(c)). The more conservative setting $(\alpha, N_t^{\text{cert}}) = (0.02, 24000)$ uses twelve times as many certification rollouts per candidate as the default $(0.10, 2000)$, without a clear improvement in the final feasible objective. On this benchmark, the additional certification effort supports a lower risk tolerance while leaving optimization performance largely unchanged.