Causal Bayesian Optimization via Causal Bandits: Open Questions

Arpan Mukherjee[1] **Zirui Yan**[2] **Ali Tajer**[2]

[1]Imperial College London, South Kensington, London, UK
[2]Rensselaer Polytechnic Institute, Troy, New York, USA,

Abstract

Causal bandits (CBs) provide a principled approach to the sequential design of interventions (experiments) on causal systems to optimize any desired utility of interest. Implicitly, CB algorithms learn the causal model while performing experiments, to the extent that the model's information is needed for optimizing the desired utility. This paper provides a perspective and framework for viewing and analyzing causal Bayesian optimization (CBO) through the lens of CBs. Despite their distinctions, CBO and CB problems share the same principle: causal structure can be used to share information across interventions and avoid treating actions as independent alternatives. Specifically, this paper first compares the two lines of literature on CBs and CBOs with respect to the roles of graph structures, intervention models, and function-class assumptions, distinguishing graph-theoretic sources of complexity from those arising from posterior surrogate-model complexity. Finally, it identifies open problems in CBO that can be addressed using recent advances in CBs.

1 INTRODUCTION

Large-scale physical and engineered systems are often modeled as networks of interconnected components whose behaviors are coupled through complex dependencies. Unlike standalone input-output systems, where each component can be studied largely in isolation, networked systems exhibit interactions in which the state or behavior of one component mediates the behavior of several others. As a result, the performance of the overall system cannot be reliably predicted by analyzing individual components separately; it must instead be understood through the behavior of the network in unison.

A natural way to represent such dependencies is through **causation**: a formal way of modeling how changes to one component of a system propagate to others. Causal relationships go beyond statistical associations by specifying how the system would respond to deliberate modifications. In a power network, for instance, causation captures how a line outage affects downstream load redistribution; in a robotic pipeline, it describes how changing a controller gain affects motion, stability, and task completion; in a biological network, it characterizes how an intervention on a gene modifies downstream expression patterns. Thus, causal relationships provide an operational framework for reasoning about the cause-and-effect relationships in complex systems.

In many real-world systems, a key goal is to optimize a network-level utility generated by such causal systems. Without knowledge of the causal relationships, this requires exploring a broad range of possible changes to the system. In contrast, should the underlying causal mechanisms be known, these could be exploited to improve learning and decision-making: one could use such structural information to narrow down the set of potentially optimal changes by leveraging information shared across changes, thereby reducing the number of samples needed to obtain a given guarantee. However, in practice, causal relationships are often unknown a priori. The learner may know part of the graph or need to estimate local mechanisms from limited data while optimizing its utility. This raises a fundamental sequential decision-making question: *how should data be allocated between learning the causal system and using the learned structure to optimize the network-level utility?*

Causal bandits. Addressing the question above involves balancing actions that are in spirit in line with the exploration–exploitation dichotomy. This, naturally, motivates the use of the multi-armed bandit (MAB) framework to resolve this trade-off. The study in Lattimore et al. [2016] introduced *causal bandits* (CB), in which the arms correspond to distinct possible *intervention* on one or more components of a causal system. An intervention's impact is inducing changes in the statistical models that generate the data of the nodes intervened. The system's utility is conventionally represented by the value of a leaf node in the causal graph (or by a dummy node added as a leaf to the graph). The learner's objective is to identify, or repeatedly select, an *optimal intervention* that maximizes the expected reward under the induced post-intervention distribution. Induced by the underlying causal structures, the arm choices and associated utilities are not statistically independent and are not independent alternatives.

Reducing causal bandits to classical MABs, where no causal information is used, and each arm's utility is estimated independently, leads to a curse of dimensionality in the cardinality of the intervention space. For instance, when the causal graph has N nodes, and there is only one possible intervention model per node, the intervention space has cardinality 2^N. With multi-valued or continuous interventions, the size of the action space can become infinite. Since canonical bandit algorithms typically incur regret that scales with the number of arms, treating interventions as unrelated actions leads to at least exponential dependence on the number of nodes N. This motivates the central question of causal bandits: *can the causal structure be learned and leveraged to share information across interventions and thereby reduce the statistical cost of sequential experimentation?*

Causal Bayesian optimization. A closely related viewpoint arises from Bayesian optimization (BO), another sequential framework for optimizing an unknown objective using adaptively chosen queries. While various MAB formulations investigate the cumulative cost of repeatedly selecting suboptimal actions, BO is often used when each evaluation is costly and the learner therefore maintains a probabilistic surrogate of the unknown objective to decide where to query next. This naturally raises a causal analog: *what if the queries are not ordinary inputs, but interventions on variables generated by a structured causal system?*

Addressing this question leads to considering *causal Bayesian optimization* (CBO) Aglietti et al. [2020]. In CBO, the learner chooses both where to intervene in the causal system and how to set the intervened variables, aiming to optimize the expected downstream outcome of the resulting intervention. Thus, CBO may be thought of as a *structured causal bandit problem* and viewed through a Bayesian lens, in which the utility of each intervention is represented through a posterior model over interventional response functions. The causal graph couples these response functions: observations collected under one intervention can inform beliefs about other interventions, enabling more sample-efficient optimization than treating each intervention independently.

Related works. The core idea of leveraging the causal structure to reduce the intervention search space has led to a growing literature that can be organized along two broad axes: i) what is known about the causal graph, and ii) what type of interventions and mechanisms are assumed. In the *known-topology* setting, early works study atomic *do*-interventions with unknown interventional distributions, using the graph to propagate information across interventions rather than treating them independently [Lu et al., 2020, Yabe et al., 2018, Xiong and Chen, 2023]. Subsequent works consider richer intervention models, including soft interventions in linear structural equation models, in which regret depends on graph parameters such as in-degree and path length rather than the full intervention count Varici et al. [2023]. More recent results extend this viewpoint to nonlinear causal mechanisms and broader classes of interventions Yan et al. [2024c]. A parallel line of work studies partially or fully unknown graphs, where the central challenge is to learn only the reward-relevant structural information needed for effective intervention selection Lu et al. [2021], Konobeev et al. [2025], Feng and Chen [2023], Elahi et al. [2024], Yan and Tajer [2024].

Contributions. This paper develops a unified view of CBO through the lens of structured causal bandits. First, we formulate CBO as a sequential intervention design problem in which interventions serve as actions and the downstream outcome determines the reward. Second, we compare the CBO and CB literature within a common framework, separating dependencies that arise from the causal graph, such as degree, path length, and intervention cardinality, from those induced by the surrogate function class, such as Gaussian-process information gain or more general function-class complexity. Finally, we identify open CBO problems that can be investigated by recent advances in causal bandits.

2 CAUSAL BANDITS

Consider a system described by a collection of random variables $X := (X_1, \ldots, X_N)$, where each variable represents one component of the networked system. The causal relationships among these variables are encoded by a directed acyclic graph (DAG) $\mathcal{G} := ([N], E)$, where an edge $i \to j$ indicates that X_i is a direct cause of X_j. In a causal Bayesian network, the observational distribution factorizes according to the graph as

$$p(x) = \prod_{i=1}^{N} p_i(x_i \mid x_{\mathrm{pa}(i)}) , \tag{1}$$

where $\mathrm{pa}(i)$ denotes the set of parents of node i. Equivalently, one may view the system through a structural equation model (SEM) as follows.

$$X_i = f_i(X_{\mathrm{pa}(i)}, U_i) , \qquad i \in [N] , \tag{2}$$

where U_i is an exogenous noise variable and f_i is the local causal mechanism generating X_i from its direct causes.

Throughout, we denote the reward variable by $Y = X_N$ and assume, following the standard causal bandit convention, that the reward node is a leaf of the graph. This is without loss of generality for decision-making, since variables downstream of the reward do not affect the learner's objective.

An intervention is an action that modifies one or more local causal mechanisms. Let $\mathcal{A}$ denote the set of available interventions. Each intervention $a \in \mathcal{A}$ targets a subset $I_a \subseteq [N]$ and replaces the original conditional kernel $p_i(x_i \mid x_{\mathrm{pa}(i)})$ by an intervention-specific kernel $q_i^a(x_i \mid x_{\mathrm{pa}(i)})$ for every $i \in I_a$. The resulting post-intervention distribution is

$$p^a(x) = \prod_{i \notin I_a} p_i(x_i \mid x_{\mathrm{pa}(i)}) \prod_{i \in I_a} q_i^a(x_i \mid x_{\mathrm{pa}(i)}) \,. \tag{3}$$

This formulation includes several common intervention classes. A *do*-intervention fixes a variable to a prescribed value, so that $q_i^a(x_i \mid x_{\mathrm{pa}(i)}) = \mathbb{1}\{x_i = x_i^a\}$. A stochastic hard intervention removes parental dependence but randomizes the target variable according to $q_i^a(x_i)$. A soft intervention is more general: it modifies the local mechanism while retaining possible dependence on the parents through $q_i^a(x_i \mid x_{\mathrm{pa}(i)})$.

The causal bandit problem proceeds sequentially. At each round $t = 1, \ldots, T$, the learner chooses an intervention $a_t \in \mathcal{A}$ based on the history and observes a realization of the network utility Y_t together with the graph feedback X_t. For an intervention $a \in \mathcal{A}$, we define its mean reward as the expectation over the utility $\mu_a := \mathbb{E}_{p^a}[Y]$. We denote an *optimal intervention* by $a^\star \in \arg\max_{a \in \mathcal{A}} \mu_a$, and seek to minimize average cumulative regret:

$$\mathfrak{R}_T := \mathbb{E}\left[\sum_{t=1}^{T} \big(\mu_{a^\star} - \mu_{a_t}\big)\right] \,. \tag{4}$$

The central question is therefore how the causal structure in (1)–(3) can be used to control regret without enumerating interventions as independent arms. The aim is to replace dependence on the intervention cardinality by dependence on structural quantities of the causal model. We note that other metrics, such as simple regret for best arm identification, may also be used.

3 CAUSAL BAYESIAN OPTIMIZATION

Bayesian optimization (BO). Let $\mathcal{X}$ be a compact decision domain and let $f : \mathcal{X} \to \mathbb{R}$ be an unknown objective function. In BO, the learner sequentially places queries $x_t \in \mathcal{X}$ for $t \in [T]$ and gleans observations

$$y_t = f(x_t) + \varepsilon_t \,, \tag{5}$$

where $\{\varepsilon_t\}_{t \geq 1}$ denotes observation noise. After t evaluations, let us denote the dataset $\mathcal{D}_t = \{(x_i, y_i)\}_{i=1}^t$. The goal is to identify a minimizer $x^\star \in \operatorname{argmin}_{x \in \mathcal{X}} f(x)$. Canonically, we assume that f belongs to a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$ associated with a positive-definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Equivalently, in the Bayesian surrogate view, one places a Gaussian process prior on f, $f \sim \mathcal{GP}(m, k)$, where $\mathcal{GP}(m, k)$ denotes a stochastic process whose finite collections are jointly Gaussian with mean function m and covariance kernel k. Conditioning on $\mathcal{D}_t$ yields the posterior process $f \mid \mathcal{D}_t \sim \mathcal{GP}(\mu_t, k_t)$, where $\mu_t : \mathcal{X} \to \mathbb{R}$ is the posterior mean and $k_t : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is the posterior covariance kernel. Accordingly, the posterior variance at x is given by $\sigma_t^2(x) = k_t(x, x)$.

Causal Bayesian optimization (CBO). Causal Bayesian optimization (CBO) studies sequential optimization over interventions in a causal Bayesian network Aglietti et al. [2020]. We keep the notation of the causal bandit setting: $X = (X_1, \ldots, X_N)$, $Y = X_N$, and $\mathcal{G} = ([N], E)$. For an intervention $a \in \mathcal{A}$, recall that p^a denotes the post-intervention distribution and denote its mean reward $\mu_a := \mathbb{E}_{p^a}[Y]$. Defining an optimal intervention $a^\star \in \arg\max_{a \in \mathcal{A}} \mu_a$, the objective in CBO is to design a sequence of interventions with minimal cumulative regret, as defined in (4). The distinguishing feature of this setting, relative to the broader causal-bandit formulation, is the regularity assumption on the local mechanisms: each f_i is assumed to belong to an RKHS $\mathcal{H}_i$, thereby enabling posterior uncertainty quantification through kernel or Gaussian-process surrogate models.

Thus, CBO can be viewed as a structured BO problem in which the objective is not a single black-box function of ordinary inputs, but an interventional response induced by a causal system. Model-based CBO Sussex et al. [2023a] makes this structure explicit by placing surrogate (Gaussian process) models on the local mechanisms $\{f_i\}_{i=1}^N$.

4 FROM CB TO CBO: ADVANCES & GAPS

Algorithmic viewpoints. CB algorithms are built on the principle that interventions are coupled through a common causal system and should not be learned as independent arms. Under do interventions, the graph is typically used to identify reward-relevant variables, prune the effective intervention set, and estimate intervention values through shared observational or interventional distributions. This yields UCB, Thompson sampling and pure-exploration algorithms whose guarantees scale with graph-dependent quantities rather than only with the number of arms [Bareinboim et al., 2015, Lattimore et al., 2016, Lu et al., 2020, Yabe et al., 2018, Xiong and Chen, 2023, Feng and Chen, 2023, Sawarni et al., 2023]. Under soft interventions, the emphasis shifts from pruning intervention targets to learning local mechanisms and propagating their uncertainty through the graph

[Sen et al., 2017, Varici et al., 2023, Yan et al., 2024a,b,c].

CBO follows the same structural principle, but implements it through Bayesian optimization. Instead of constructing bandit rules directly over arms, CBO places probabilistic surrogates on interventional response surfaces and selects interventions using acquisition functions. Early CBO methods exploit known causal structure to restrict the intervention search space, in a spirit related to possibly optimal minimal intervention sets [Aglietti et al., 2020, Gultchin et al., 2023]. Model-based CBO goes further by placing GP surrogates on the local causal mechanisms and propagating posterior uncertainty through the graph, paralleling the soft-intervention view in causal bandits [Sussex et al., 2023a, Varici et al., 2023]. Thus, causal bandits and CBO share the same organizing idea: structural information sharing across interventions. At the same time, they also differ in how uncertainty is represented and operationalized: causal bandits through regret-minimizing sampling rules, and CBO through surrogate models and acquisition functions.

Existing CBO guarantees. Aglietti et al. [2020] introduced causal global optimization and the CBO algorithm, combining causal Gaussian-process priors, causal expected improvement, and an ϵ-greedy rule for the observation–intervention trade-off. Their analysis relies on identifiability through do intervention and the usual properties of the surrogate model, but it does not provide a finite-time regret guarantee for the full CBO procedure. Subsequent investigations extend this design in several directions: Aglietti et al. [2021] extend CBO to dynamic settings with time-dependent causal graph evolution, Aglietti et al. [2023] incorporate feasibility constraints through constrained causal acquisition functions, Arsenyan et al. [2023] study contextual CBO, where observational variables are used as contexts for choosing intervention values, and the algorithm searches over policy scopes that combine contextual and interventional variables, Gultchin et al. [2023] extend CBO to functional contextual interventions, and Bhatija et al. [2025] extend the framework to multi-objective causal optimization through Pareto-style intervention pruning. For unknown graph structure, Branchini et al. [2023] incorporate graph uncertainty through an information-theoretic acquisition rule, while Durand et al. [2025] focus on learning a posterior over the direct parents of the target; this latter strategy shares a similar local-structure-learning idea with unknown-graph causal bandits Lu et al. [2021], Konobeev et al. [2025]. Sussex et al. [2023a] formalize model-based CBO (MCBO), which learns the structural mechanisms f_i rather than fitting a separate GP to each intervention-response surface. This is close to CB algorithms that estimate reusable local quantities rather than the full arm means. Under a known graph, full observation of the system, calibrated GP uncertainty, and smooth soft interventions, MCBO obtains the first non-asymptotic sublinear cumulative-regret bounds for CBO of the order

$$\tilde{\mathcal{O}}(d^L \gamma_T^{\frac{L+1}{2}} N\sqrt{T}) \,, \tag{6}$$

where d is the maximum degree, L is the longest causal path and γ_T is the information gain. Thus, the guarantee avoids direct dependence on the number of intervention sets and instead depends on graph-specific parameters and information gain. Sussex et al. [2023b] extend this model-based perspective to adversarial CBO, where uncontrolled external interventions are handled through multiplicative weights and optimistic counterfactual reward estimates.

Advances in CB. The CB literature has developed sharper finite-time analyses for related, but not identical, intervention models. For known graphs and linear SEMs with soft interventions, Varici et al. [2023] avoid estimating the 2^N interventional reward distributions directly and instead, estimate the SEM parameters. Their upper bound scales as

$$\tilde{\mathcal{O}}\left(d^{L-\frac{1}{2}}\sqrt{NT}\right), \tag{7}$$

with a lower bound

$$\Omega\left(d^{\frac{L}{2}-2}\sqrt{T}\right). \tag{8}$$

Yan et al. [2024c] extend this mechanism-estimation viewpoint beyond linear SEMs to general Lipschitz structural functions and generalized soft interventions, including infinitely many interventions. Their regret bound depends on the graph parameters and on the complexity of the mechanism class, through quantities such as the eluder dimension $\dim(\mathcal{F})$ and covering number $\mathrm{cn}(\mathcal{F})$ as

$$\tilde{\mathcal{O}}\Big(Kd^{L-1}\sqrt{T\,\dim(\mathcal{F})\log \mathrm{cn}(\mathcal{F})}\Big), \tag{9}$$

with lower bounds specialized to several SEM classes.

Gaps. Recent advances in causal bandits point to several open directions for CBO. First, existing performance guarantees should be sharpened to separate intrinsic graph dependence from artifacts of GP confidence propagation. In particular, a more refined theory should clarify how quantities such as graph depth, maximum in-degree, reward ancestors, and graph cardinality interact with information-gain terms from the surrogate model. Second, the current model-based CBO largely assumes known graphs and fully observed systems; extending CBO to unknown or partially known topologies, latent confounding, and partial observation remains a major open challenge, with recent causal-bandit developments offering natural starting points [Lee and Bareinboim, 2018, Yan et al., 2024b, Park et al., 2025, Elahi et al., 2024, Yan and Tajer, 2024]. Finally, computation is a central bottleneck: optimistic CBO acquisition rules may require global optimization over intervention parameters and plausible causal mechanisms. A mature theory of CBO should therefore pair regret guarantees with tractable algorithms, or develop approximation-aware regret bounds for practically implementable acquisition procedures.

References

Virginia Aglietti, Xiaoyu Lu, Andrei Paleyes, and Javier González. Causal bayesian optimization. In *Proc. International Conference on Artificial Intelligence and Statistics*, virtual, Augeust 2020.

Virginia Aglietti, Neil Dhir, Javier González, and Theodoros Damoulas. Dynamic causal bayesian optimization. *Proc. Advances in Neural Information Processing Systems*, 2021.

Virginia Aglietti, Alan Malek, Ira Ktena, and Silvia Chiappa. Constrained causal bayesian optimization. In *Proc. International Conference on Machine Learning*, Honolulu, HI, July 2023.

Vahan Arsenyan, Antoine Grosnit, Haitham Bou-Ammar, and Arnak Dalalyan. Contextual causal bayesian optimisation. *arXiv:2301.12412*, 2023.

Elias Bareinboim, Andrew Forney, and Judea Pearl. Bandits with unobserved confounders: A causal approach. In *Proc. Advances in Neural Information Processing Systems*, Montréal, Canada, December 2015.

Shriya Bhatija, Paul-David Zuercher, Jakob Thumm, and Thomas Bohné. Multi-objective causal bayesian optimization. *arXiv:2502.14755*, 2025.

Nicola Branchini, Virginia Aglietti, Neil Dhir, and Theodoros Damoulas. Causal entropy optimization. In *Proc. International Conference on Artificial Intelligence and Statistics*, Valencia, Spain, April 2023.

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

Muhammad Qasim Elahi, Mahsa Ghasemi, and Murat Kocaoglu. Partial structure discovery is sufficient for noregret learning in causal bandits. In *Proc. Advances in Neural Information Processing Systems*, Vancouver, Canada, December 2024.

Shi Feng and Wei Chen. Combinatorial causal bandits. In *Proc. AAAI Conference on Artificial Intelligence*, Washington, D.C., February 2023.

Limor Gultchin, Virginia Aglietti, Alexis Bellot, and Silvia Chiappa. Functional causal bayesian optimization. In *Proc. Uncertainty in Artificial Intelligence*, Pittsburgh, PA, July 2023.

Mikhail Konobeev, Jalal Etesami, and Negar Kiyavash. Causal bandits without graph learning. In *Proc. Conference on Causal Learning and Reasoning*, Lausanne, Switzerland, May 2025.

Finnian Lattimore, Tor Lattimore, and Mark D Reid. Causal bandits: Learning good interventions via causal inference. In *Proc. Advances in Neural Information Processing Systems*, Barcelona, Spain, December 2016.

Sanghack Lee and Elias Bareinboim. Structural causal bandits: Where to intervene? In *Proc. Advances in Neural Information Processing Systems*, pages 2573–2583, Montréal, Canada, December 2018.

Yangyi Lu, Amirhossein Meisami, Ambuj Tewari, and William Yan. Regret analysis of bandit problems with causal background knowledge. In *Proc. Conference on Uncertainty in Artificial Intelligence*, virtual, August 2020.

Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Causal bandits with unknown graph structure. In *Proc. Advances in Neural Information Processing Systems*, virtual, December 2021.

Min Woo Park, Andy Arditi, Elias Bareinboim, and Sanghack Lee. Structural causal bandits under markov equivalence. In *Proc. Advances in Neural Information Processing Systems*, San Diego, CA, December 2025.

Ayush Sawarni, Rahul Madhavan, Gaurav Sinha, and Siddharth Barman. Learning good interventions in causal graphs via covering. In *Proc. Conference on Uncertainty in Artificial Intelligence*, Pittsburgh, PA, 2023.

Rajat Sen, Karthikeyan Shanmugam, Alexandros G. Dimakis, and Sanjay Shakkottai. Identifying best interventions through online importance sampling. In *Proc. International Conference on Machine Learning*, Sydney, Australia, August 2017.

Scott Sussex, Anastasiia Makarova, and Andreas Krause. Model-based causal bayesian optimization. In *Proc. International Conference on Learning Representations*, Kigali, Rwanda, May 2023a.

Scott Sussex, Pier Giuseppe Sessa, Anastasia Makarova, and Andreas Krause. Adversarial causal bayesian optimization. In *Proc. International Conference on Learning Representations*, Kigali, Rwanda, May 2023b.

Burak Varici, Karthikeyan Shanmugam, Prasanna Sattigeri, and Ali Tajer. Causal bandits for linear structural equation models. *Journal of Machine Learning Research*, 24(297): 1–59, 2023.

Nuoya Xiong and Wei Chen. Combinatorial pure exploration of causal bandits. In *Proc. International Conference on Learning Representations*, Kigali, Rwanda, May 2023.

Akihiro Yabe, Daisuke Hatano, Hanna Sumita, Shinji Ito, Naonori Kakimura, Takuro Fukunaga, and Ken-ichi

Kawarabayashi. Causal bandits with propagating inference. In *Proc. International Conference on Machine Learning*, Stockholm, Sweden, July 2018.

Zirui Yan and Ali Tajer. Linear causal bandits: Unknown graph and soft interventions. In *Proc. Advances in Neural Information Processing Systems*, Vancouver, Canada, December 2024.

Zirui Yan, Arpan Mukherjee, Burak Varıcı, and Ali Tajer. Improved bound for robust causal bandits with linear models. In *Proc. IEEE International Symposium on Information Theory*, Athens, Greece, July 2024a.

Zirui Yan, Arpan Mukherjee, Burak Varıcı, and Ali Tajer. Robust causal bandits for linear models. *IEEE Journal on Selected Areas in Information Theory*, 5:78–93, March 2024b.

Zirui Yan, Dennis Wei, Dmitriy Katz-Rogozhnikov, Prasanna Sattigeri, and Ali Tajer. Causal bandits with general causal models and interventions. In *Proc. International Conference on Artificial Intelligence and Statistics*, Valencia, Spain, May 2024c.