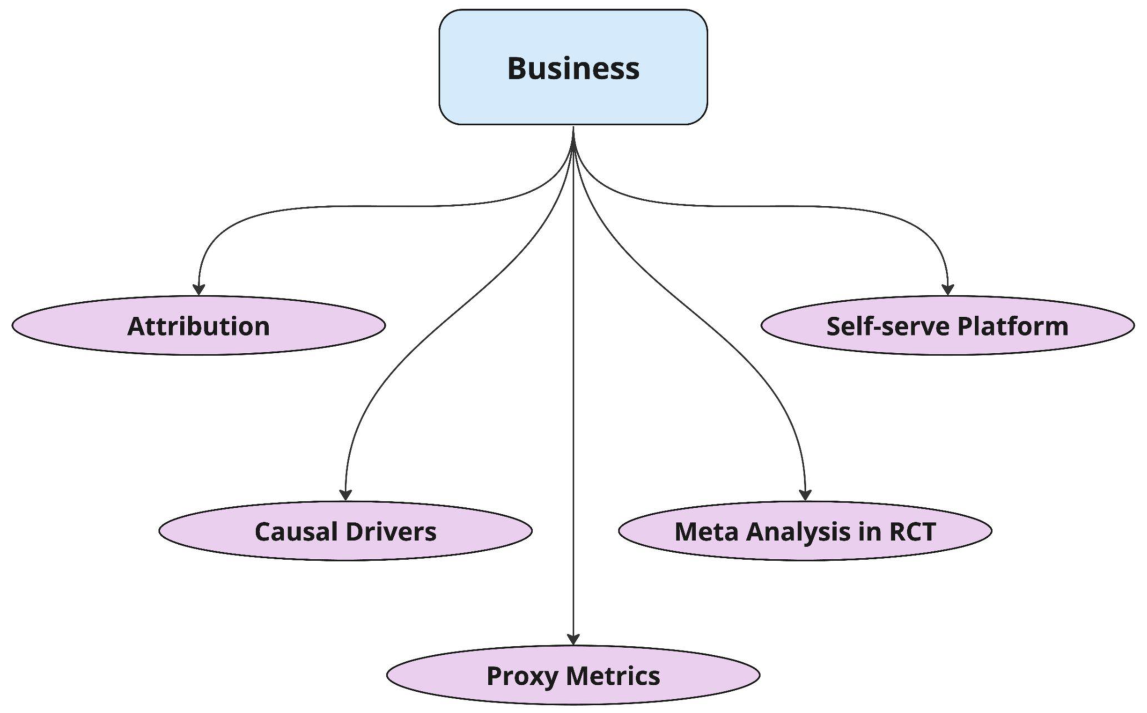
Causal Inference

Jaiprakash R, Ilari Paasivirta , Vinod Reddy

SCAN ME

Causal Inference at Scale in Dream Sports

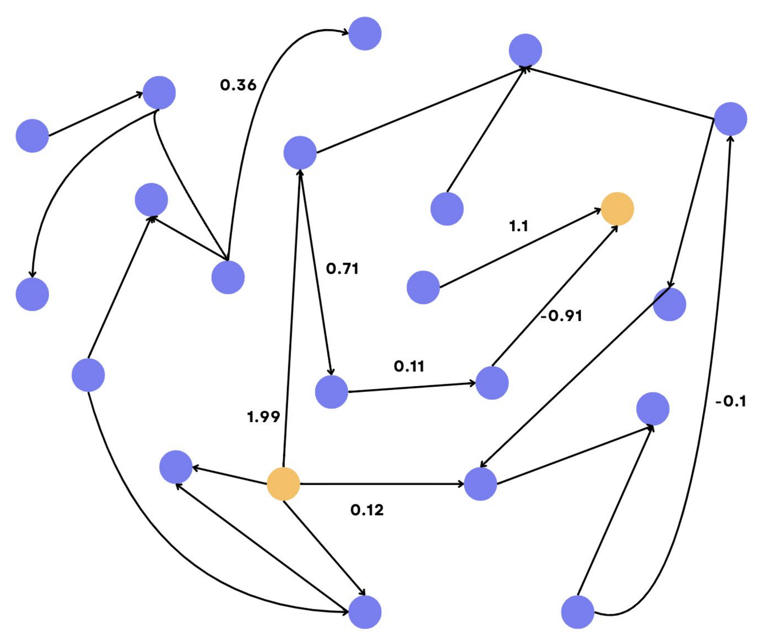


- We aim to continuously improve user engagement and business metrics using causal analysis to inform business decisions.
- We have established dozens of causal relationships, and proxy metrics.
- We built a highly versatile Causal Inference Platform that is powered by Ray and our ML Platform enabling us to build complex counterfactual models at scale.

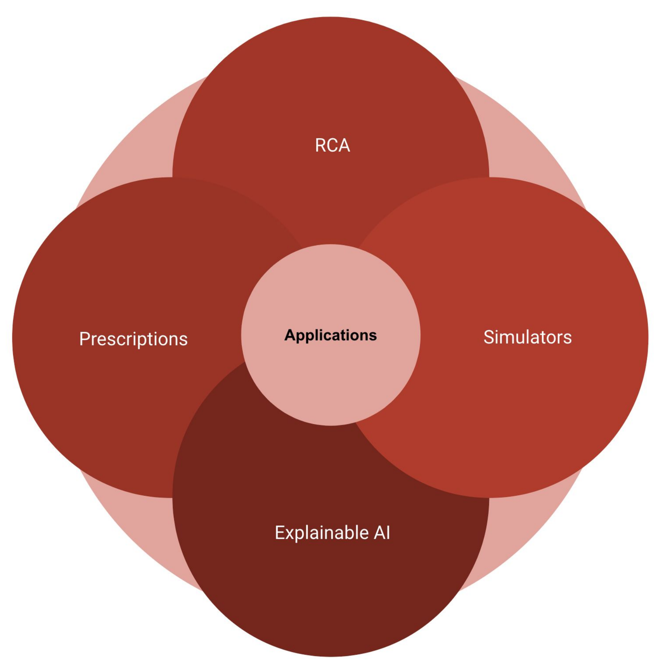
Key Area 1 - Causal Representation Learning

Building an org-wide graph that captures how key metrics are related to one another will improve decision-making in multiple different areas, such as improving automated systems, root cause analysis, and business decision-making. The key metrics range from various user engagement metrics, to typical business metrics such as revenue. Some aims include,

1. Estimating the causal graph structure between our key metrics.
2. Estimating the strength of those causal relationships.
3. How to capture the changes in causal relationships, given that in an evolving business environment we cannot assume that relationships remain static.

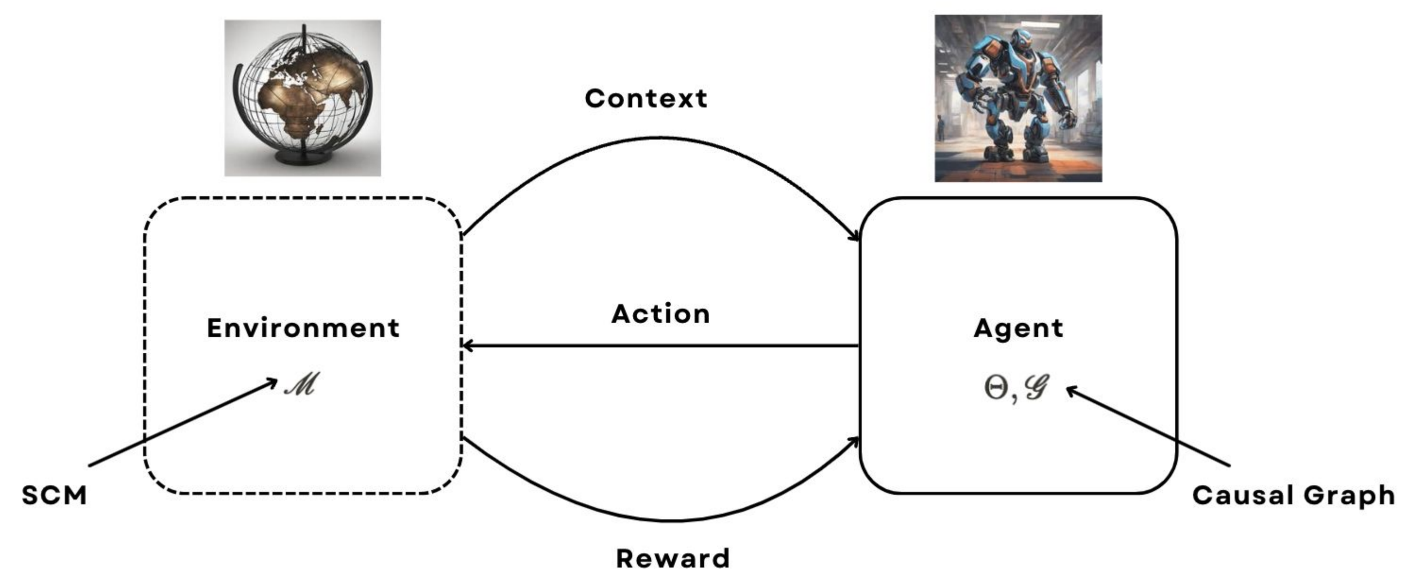


A causal graph with estimated edge strengths



Key Area 2 - Causal RL Systems

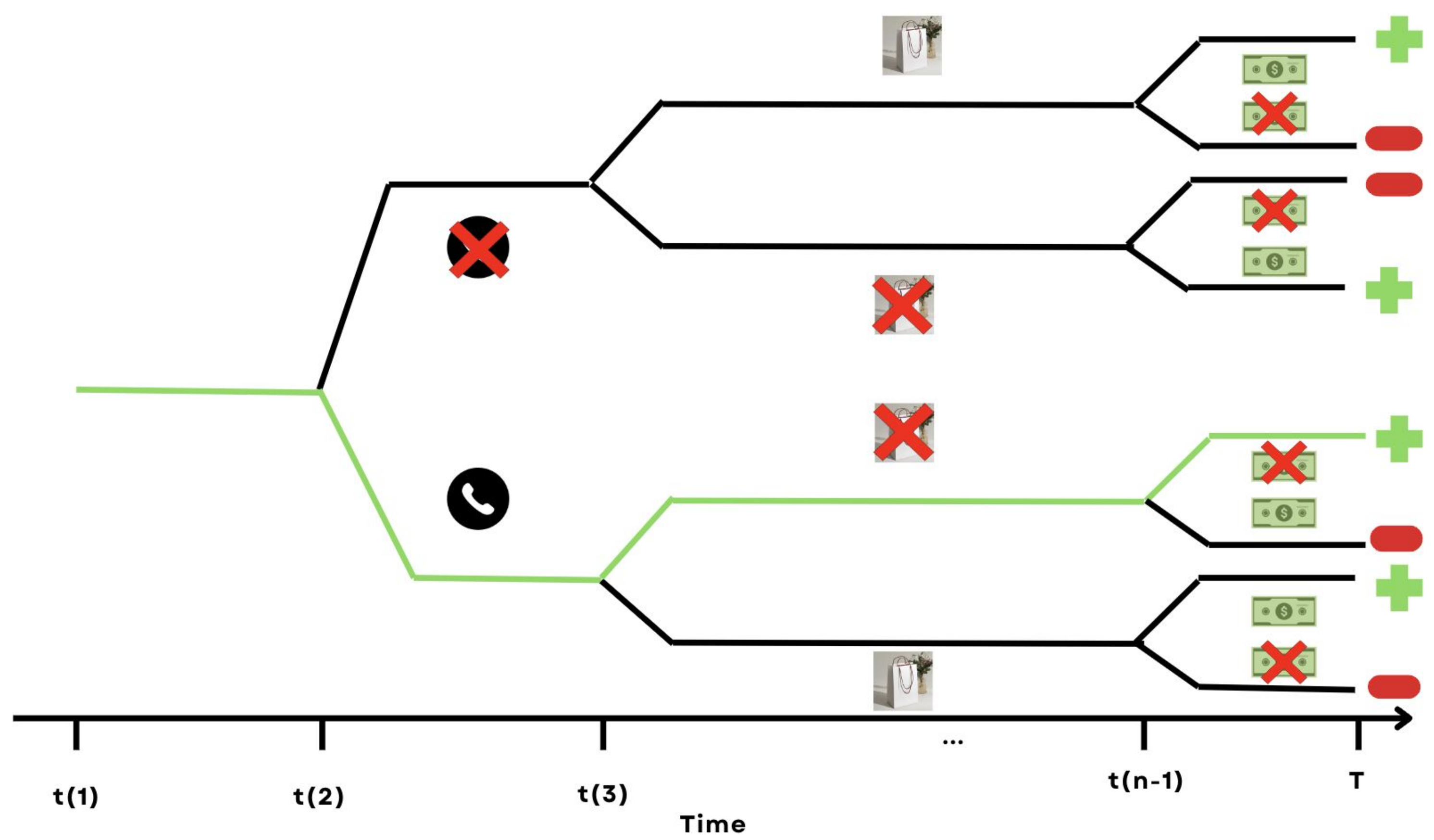
Causal RL Systems will enable policies that optimize long term outcomes driven by counterfactual learning in our systems.



Suppose $\mathscr{S}$ is the state space that encompasses user context and system state of Dream Sports at any given time t, and $\mathscr{A}$ is a high dimensional action space covering a wide range of actions such as discounts, assistance, recommendations, nudges and other offerings.

$$\text{Estimate the policy } \pi_t(\mathbf{s_{t-1}}) = \arg\max_{\mathbf{a_t}} \sum_{t=1}^{T} \gamma^t R(\mathbf{s_{t-1}}, \mathbf{a_t})$$

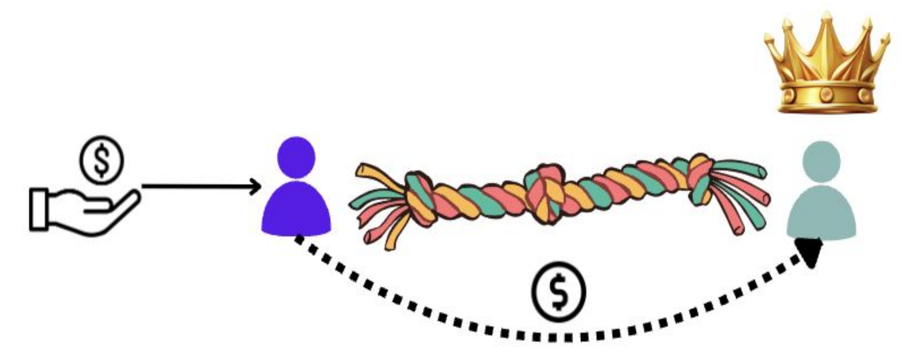
$$\text{where action } \mathbf{a_t} \in \mathscr{A} \text{ and state } \mathbf{s_{t-1}} \in \mathscr{S}$$



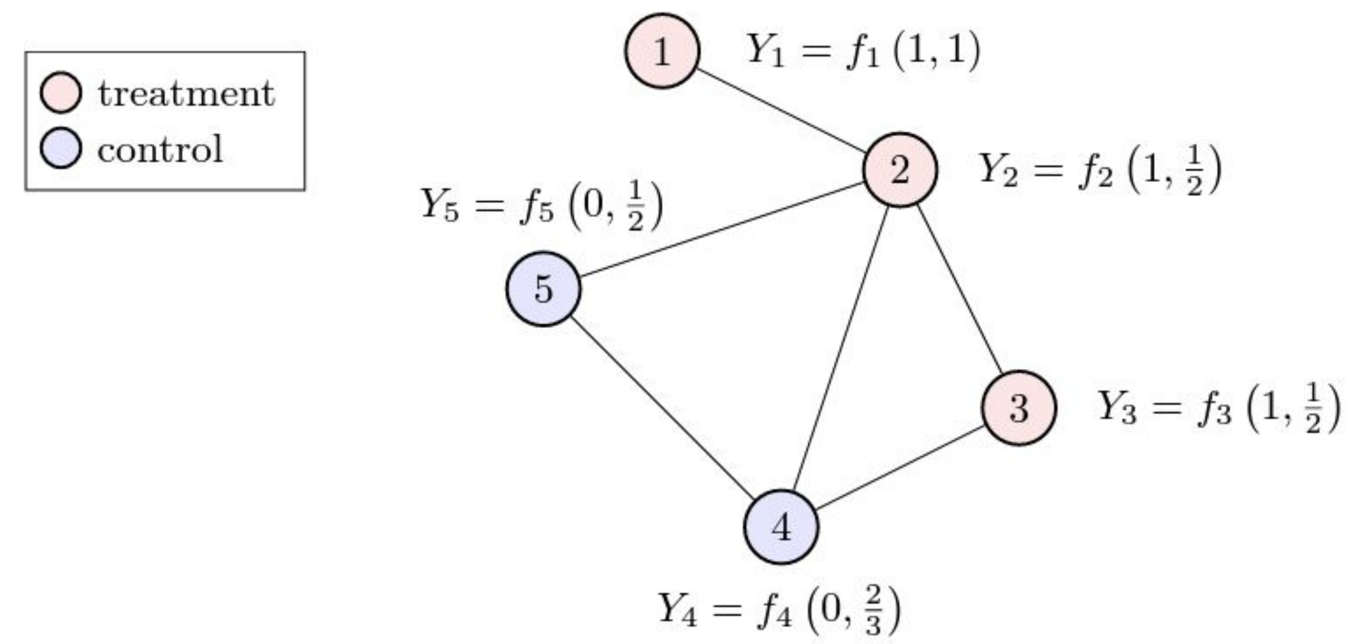
A pictorial representation of the goal : optimal sequence of decisions from a large action space.

Key Area 3 - Diffusion of Money

Ignoring spillover effects of discounts given to users leads to a biased estimate of the ROI. Quantifying these indirect effects would lead to more accurate estimation of the ROI, and thus help us optimize spending decisions.



- The discounts given to users in the Treatment Group (TG) spill over to users in the Control Group (CG) when they compete in a fantasy game and a user in the CG wins.



- In the above example graph, nodes represent users and edges represent whether users participated in the same contest.
- Outcome $Y = f(T, q)$, where T is the binary treatment status, and q is the fraction of other connected players who were treated.

In the presence of spillover effects, the observed outcomes for an individual are no longer only a function of their treatment assignment, but also their neighbours.

References

1. Hiroshi Morioka et al. (2023). "Causal Representation Learning Made Identifiable by Grouping of Observational Variables." In: preprint arXiv:2310.15709
2. Lixin Zou et al. (2019). "Reinforcement Learning to Optimize Long Term User Engagement in Recommender Systems.". In: KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, p. 2810-2818.
3. Junzhe Zhang et al. (2020). "Designing Optimal Dynamic Treatment Regimes: A Causal Reinforcement Learning Approach". In: Proceedings of the 37th International Conference on Machine Learning.
4. Shuangning Li et al. (2022). "Random Graph Asymptotics for Treatment Effect Estimation under Network Interference".In: preprint arXiv:2007.13302