Linear Causal Disentanglement via Interventions

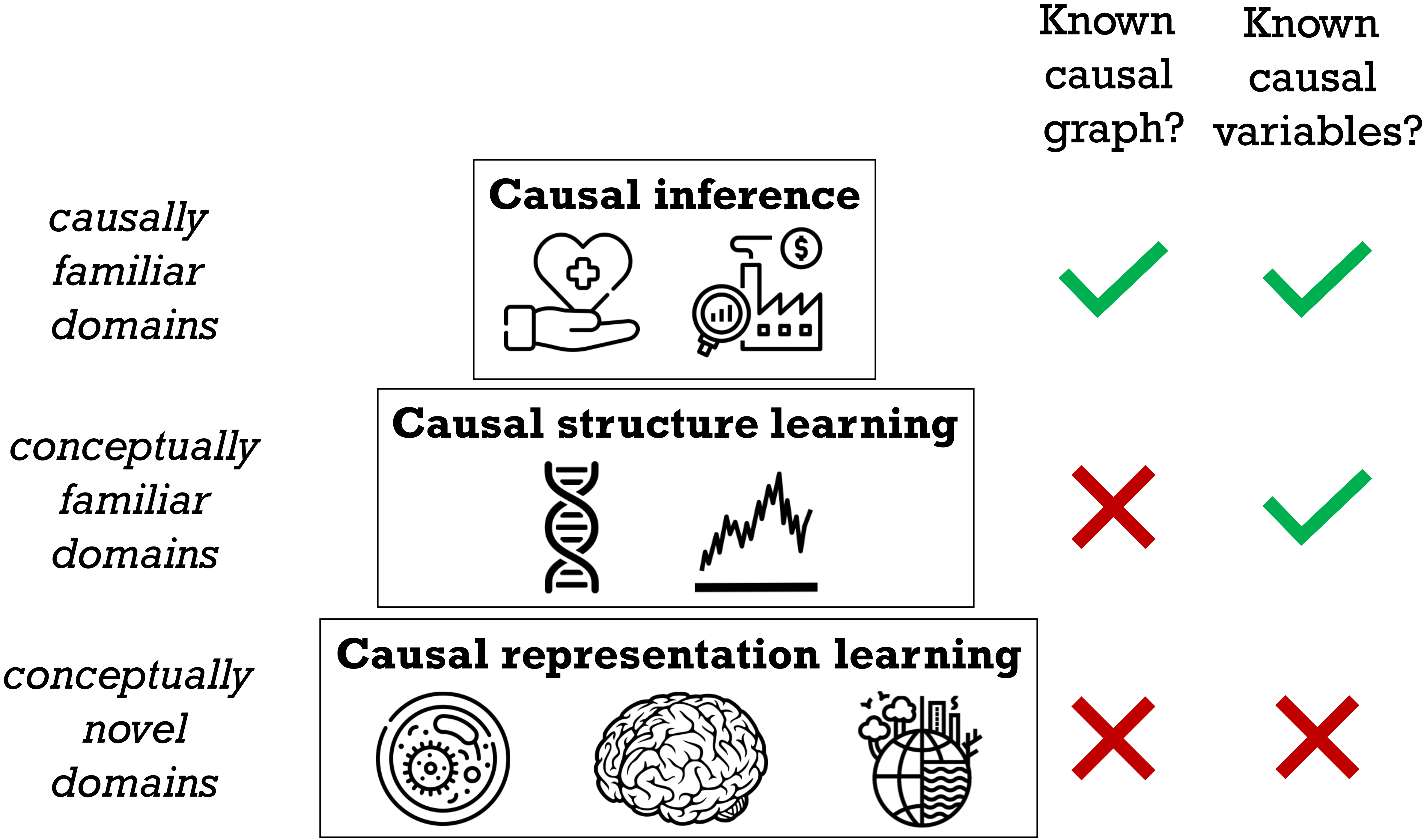
Chandler Squires*[1,2], Anna Seigal*[1,3], Salil Bhate[1],
Caroline Uhler[1,2]

[1]Broad Institute of MIT and Harvard

[2]Laboratory for Information and Decision Systems, MIT

[3]School of Engineering and Applied Science, Harvard

The context…

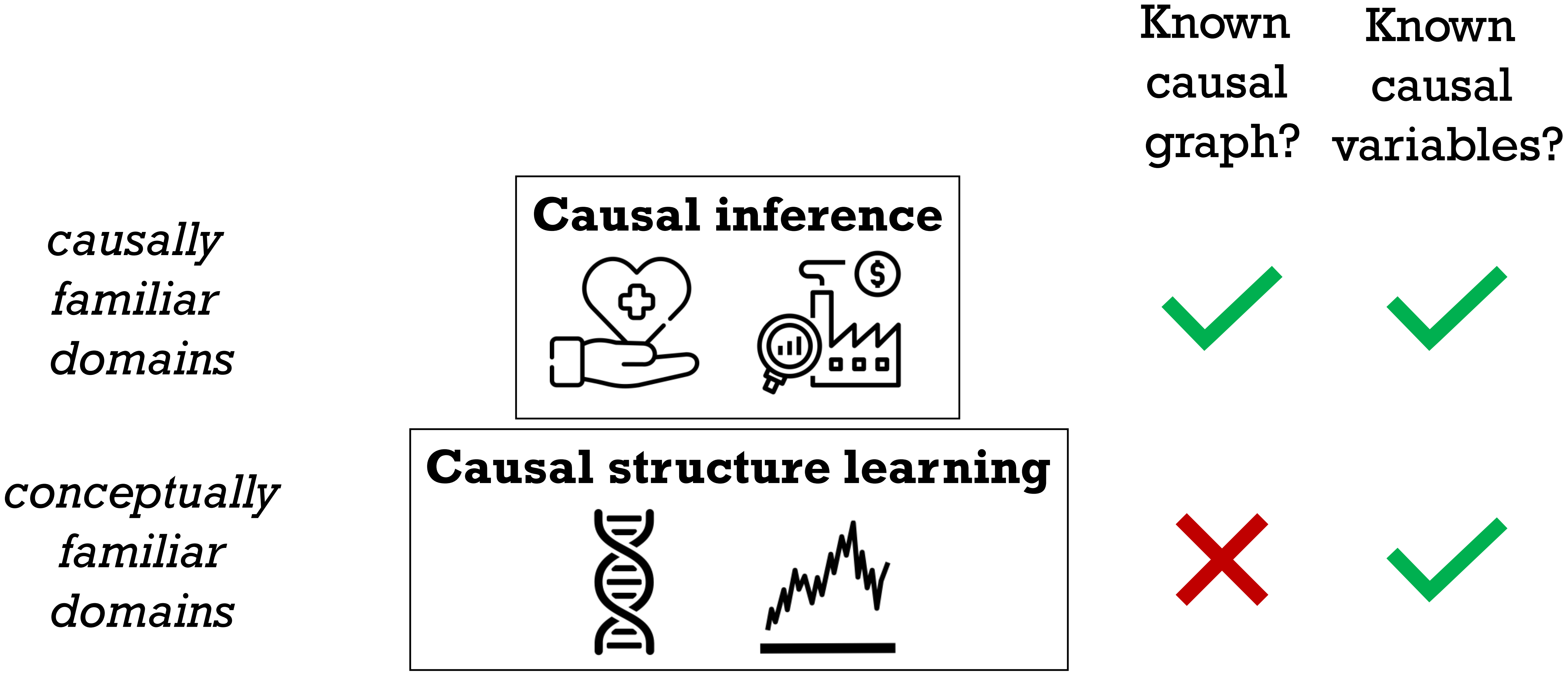

Known causal graph?
Known causal variables?
causally familiar domains
Causal inference
conceptually familiar domains
Causal structure learning
conceptually novel domains
Causal representation learning

Known causal graph?

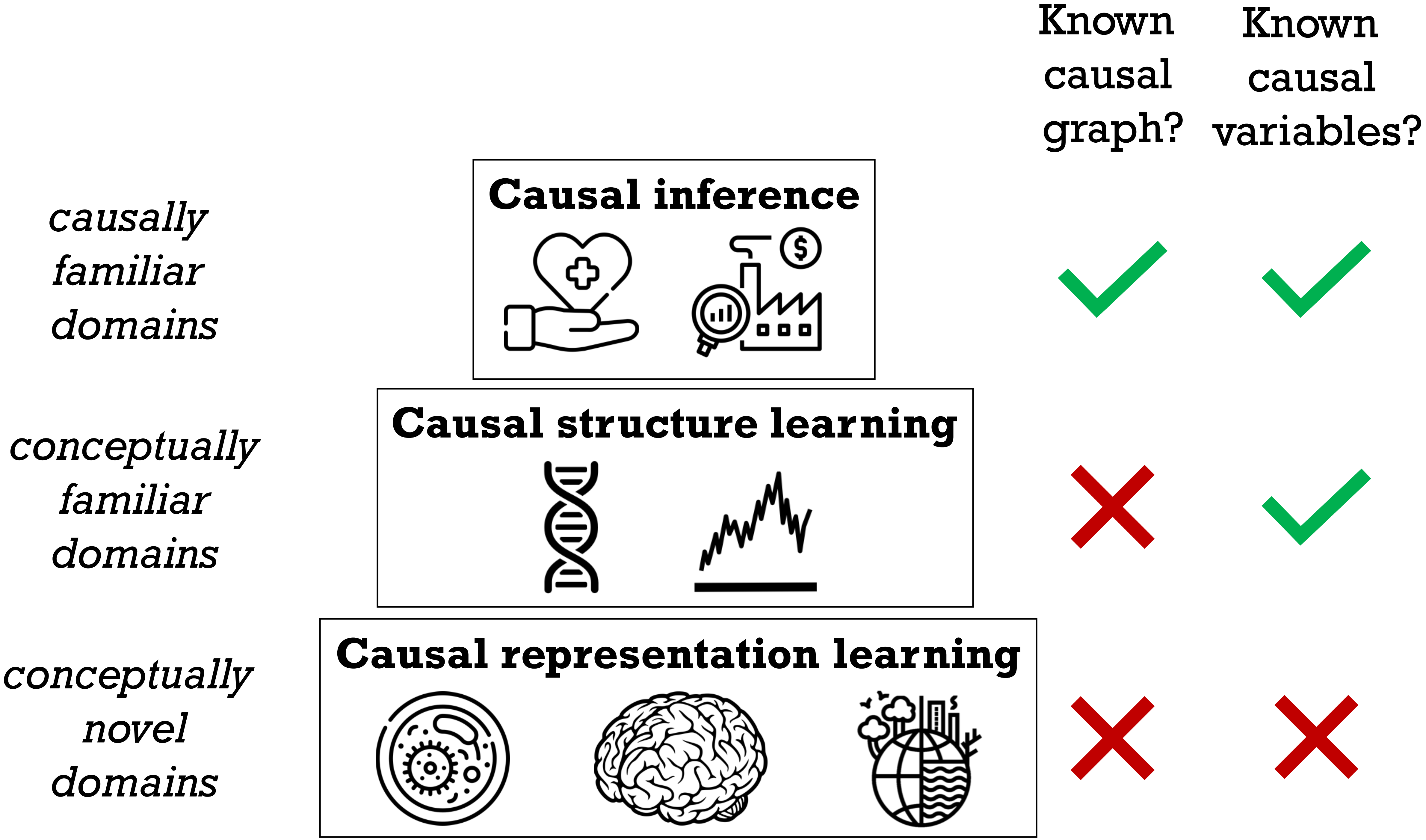
Known causal variables?

causally familiar domains

Causal inference



Known causal graph?
Known causal variables?
causally familiar domains
Causal inference
conceptually familiar domains
Causal structure learning

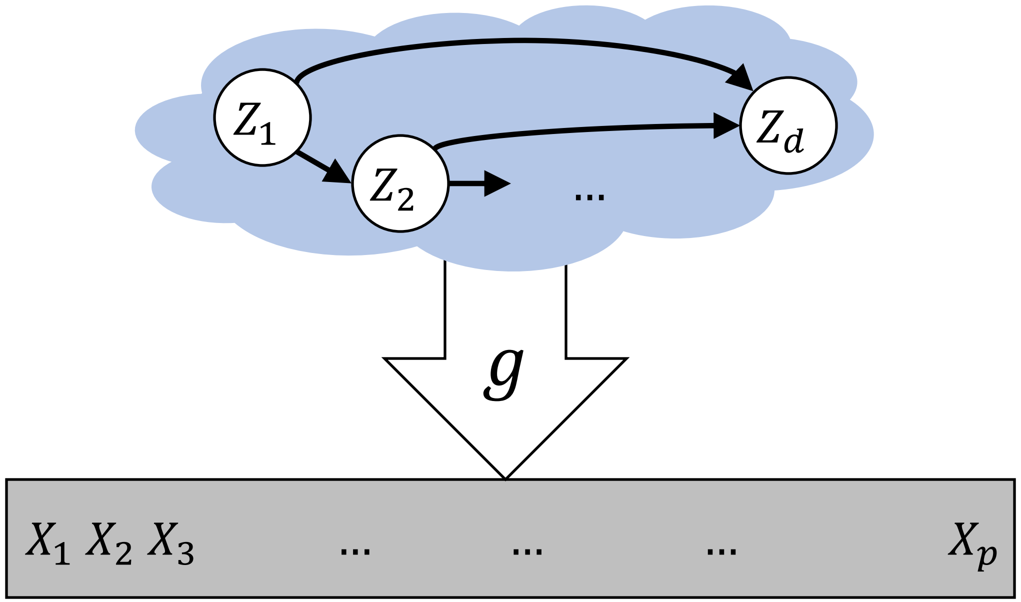

Known causal graph?
Known causal variables?
causally familiar domains
Causal inference
conceptually familiar domains
Causal structure learning
conceptually novel domains
Causal representation learning

Causal Disentanglement

Macro-variables

Mixing function

Micro-variables



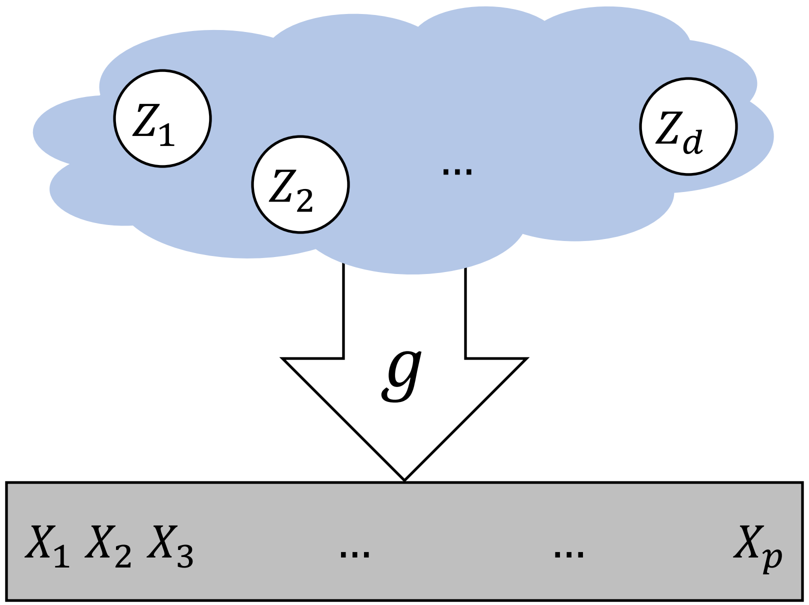
A central question:
Identifiability

Identifiability = A unique model explains the data we observe.

Approaches to identifiability the causal disentanglement problem

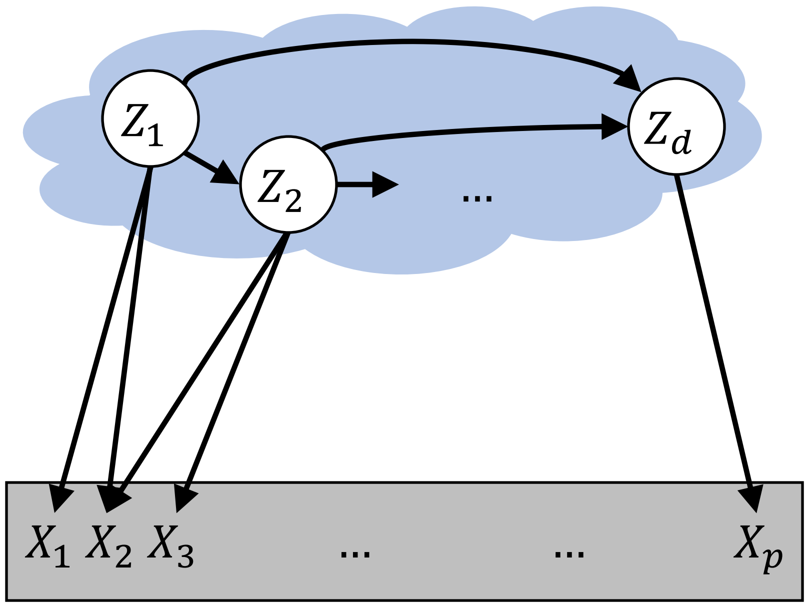
Identifiability = A unique model explains the data we observe.

Restrict latent DAG $\mathcal{G}$



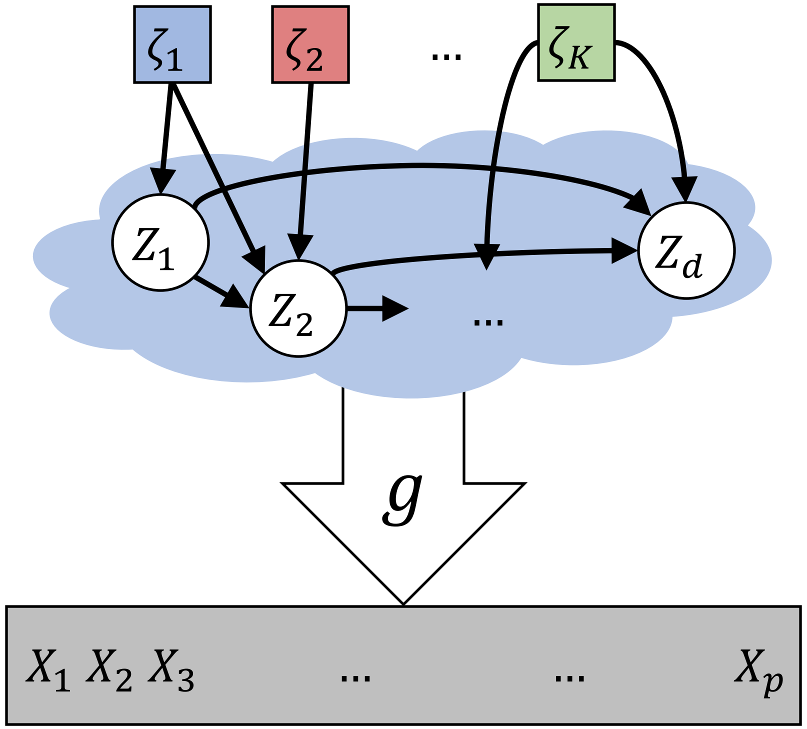
Linear ICA (Comon 1994)
Nonlinear ICA (Hyvärinen '19)

Restrict mixing function g



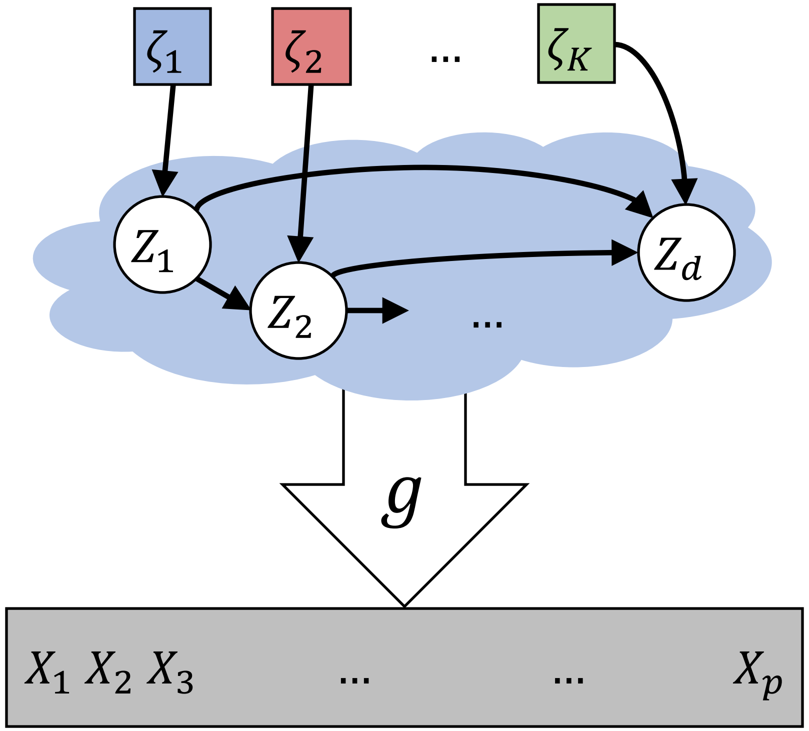
Most work on latent DAG recovery
(Silva '06, Halpern '15, Cai '19,
Kivva '21, Xie '20, Xie '22)

Learning from contexts



This work
Liu '22, Ahuja '22

Our setting



Single node interventions:
$T(k) = \{i_k\}$

Linear latent model:
$Z = B_k^{-1}\varepsilon$, ε independent,
B_k upper triangular

Linear observations:
$X = GZ$, G full column rank

Our identifiability guarantees

Perfect interventions: One intervention per latent node is sufficient, and in the worst case, necessary for identifying $(T, B_0, B_1, \dots, B_K, G)$.

Soft interventions: One intervention per latent node is sufficient, and in the worst case, necessary for identifying the latent graph up to its transitive closure.

Constructive approach

Key identity: $$\Theta_k - \Theta_0 = \underbrace{\left(H^\top B_k^\top \boldsymbol{e}_{i_k}\right)^{\otimes 2} - \left(H^\top B_0^\top \boldsymbol{e}_{i_k}\right)^{\otimes 2}}$$

Rank = 1 if i_k is a source node, rank = 2 otherwise.
Used to find (scaled) $\boldsymbol{h}_{i_k}$ when i_k is a source node.

General idea: $\mathrm{rowspan}(\Theta_k - \Theta_0) \subseteq \{\boldsymbol{h}_i : i \in \mathcal{J}\}$ if and only if $\mathcal{J} = \mathrm{pa}(i_k) \cup \{i_k\}$, so we can iteratively recover (1) the partial order over i_k's and (2) the corresponding rows of H.

arxiv.org/abs/2211.16467