

Causal Inference with Graph Neural Networks

lecture-13

Course on Graph Neural Networks (Winter Term 20/21)

Christian Medeiros Adriano (christian.adriano@hpi.de) - “Chris”

Sona Ghahremani (sona.ghahremani@hpi.de)

Prof. Dr. Holger Giese (holger.giese@hpi.de)

Quick recap – Where are we now?

1. Graph Metrics and Random Models
2. Graph Classification - Clustering
3. Graph Sampling - Random Walks
4. Graph Embeddings - Message Passing
5. PageRank
6. Graph Structure Learning
7. Graph Convolutional Networks
8. Graph Attention Networks
9. Graph Evolution Networks
10. Temporal Graph Networks
11. Graph Neural Differential Equations
12. Deep Graph Generative Models

} Description models

} Prediction models

13. Causal Graph Neural Networks

14. Propagation Graph Neural Networks
 - Network Effects, Cascading and Contagion
 - Outbreak Detection and Influence Maximization

} Intervention models

Design concerns

Understand a phenomenon
Extract features
Stablish baselines
Preprocessing data

Predict an outcome
ML architecture and pipeline
Training models
Evaluation models

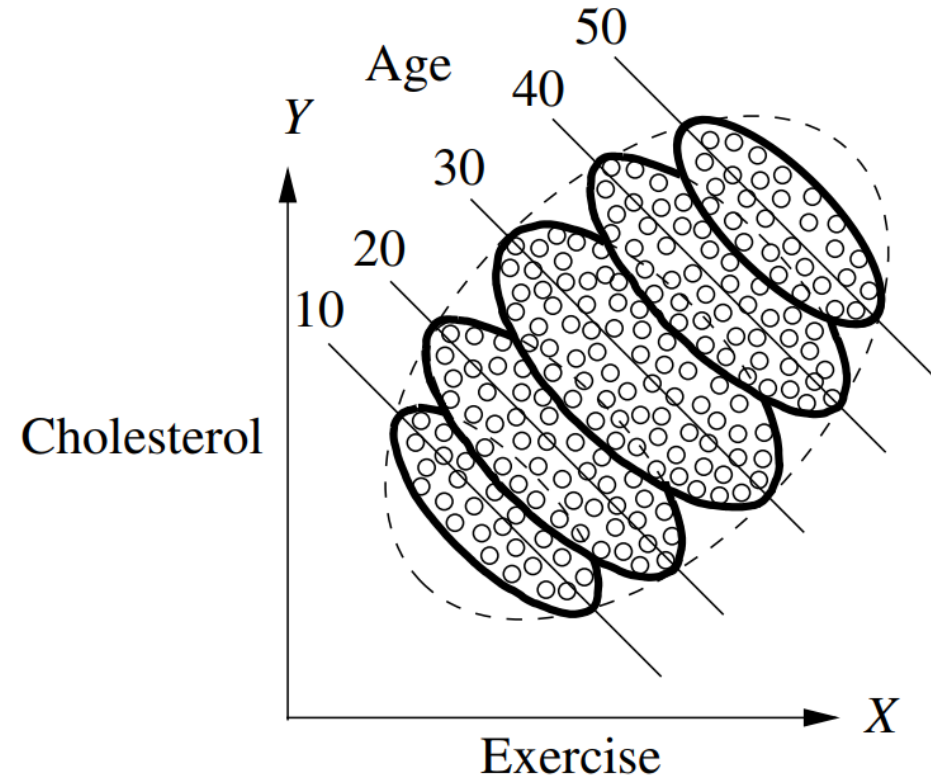
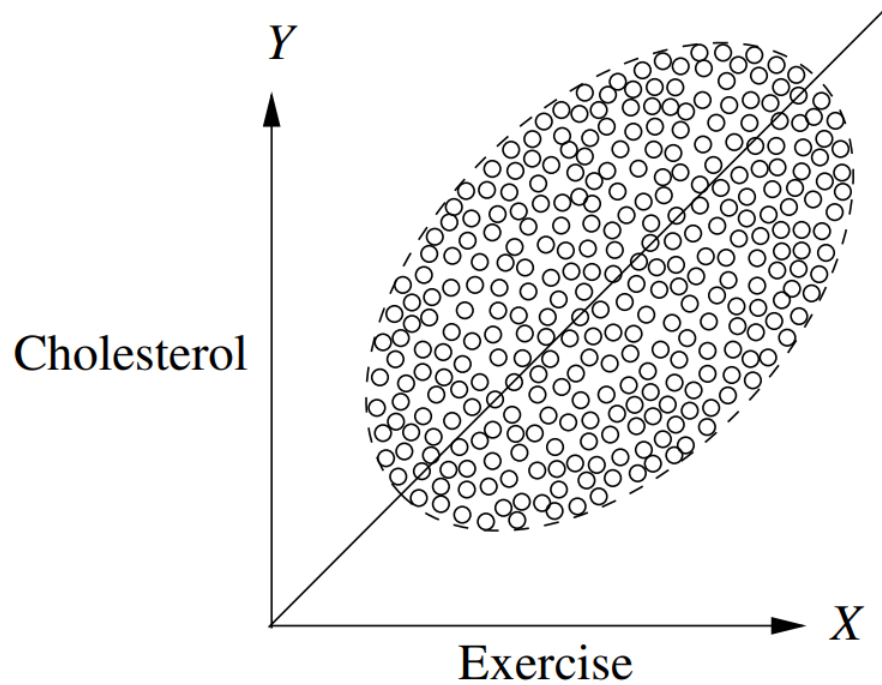
Effects of interventions
Risks of confounding
Causal structure

Definition: this paradox happens when a statistic holds for the entire population, but reverses at each subpopulation

	Overall	Patients with small stones	Patients with large stones
Treatment <i>a</i> : Open surgery	78% (273/350)	93% (81/87)	73% (192/263) more
Treatment <i>b</i> : Percutaneous nephrolithotomy	83% (289/350)	87% (234/270) more	69% (55/80)

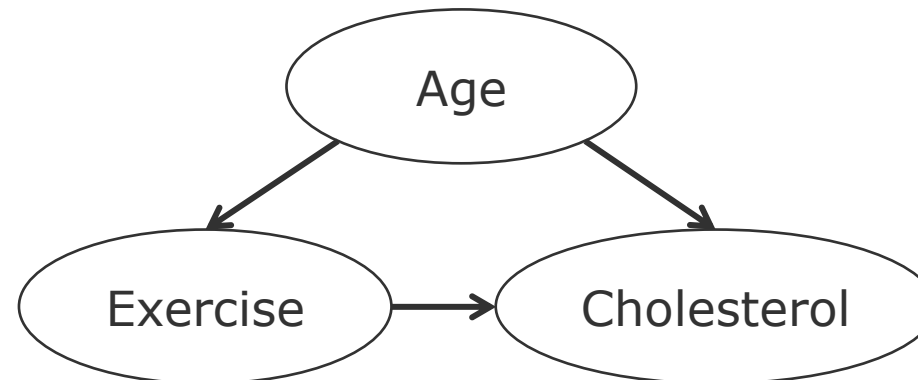
Source: [Peters et al. 2017] [Bottou et al., 2013, Charig et al., 1986, tables I and II]

Simpson Paradox - 2



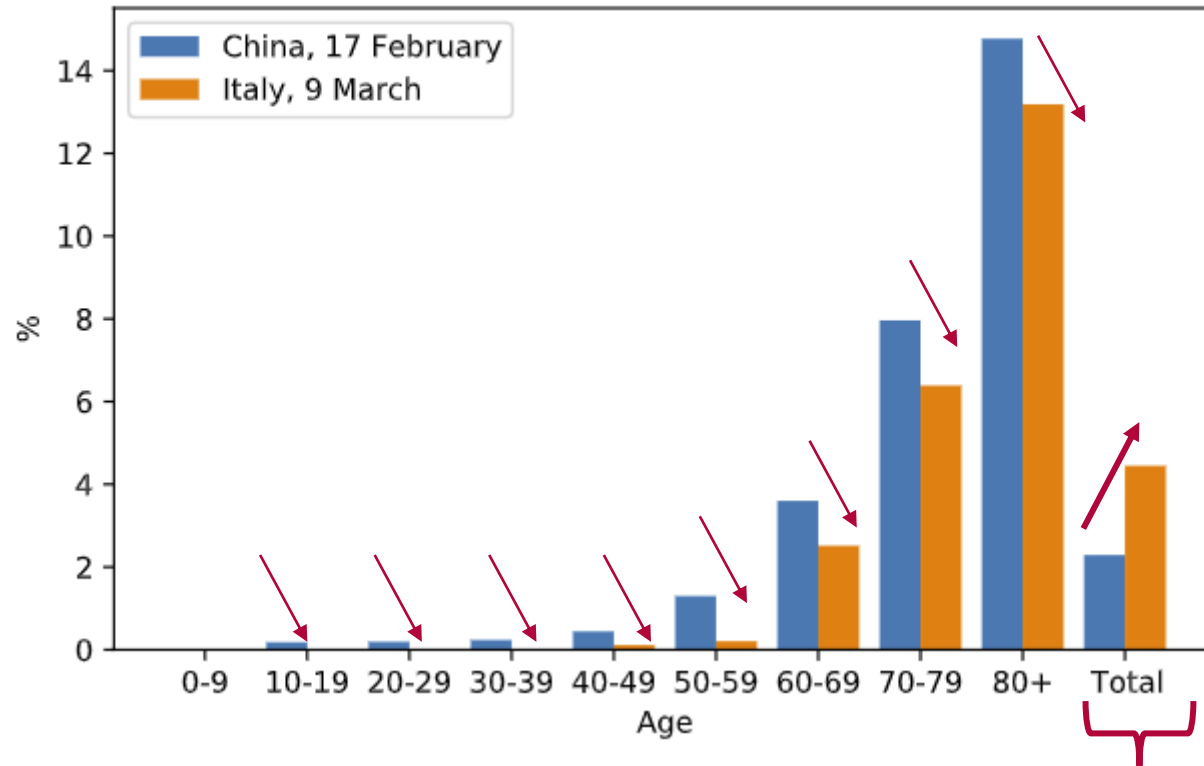
Source: Pearl, Glymour & Jewell, Causal Inference in Statistics: A Primer, 2016

Solution – condition on age!



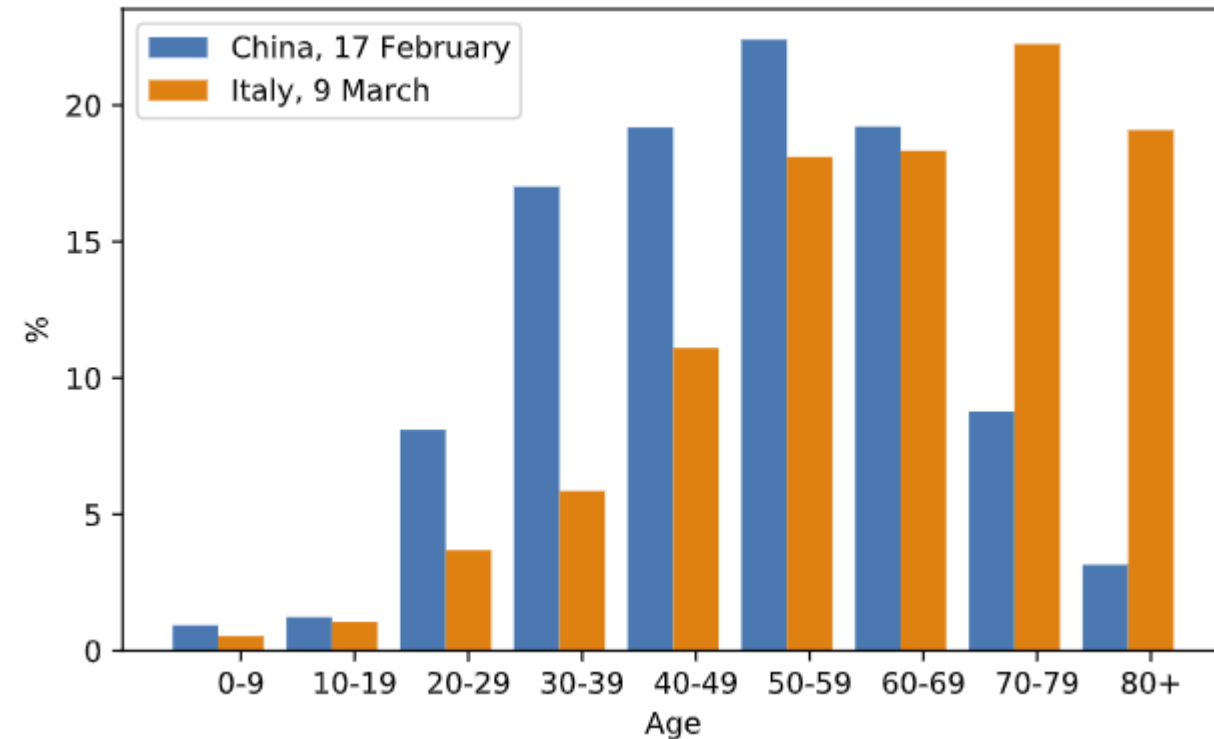
Covid-19 fatality rates China and Italy

Case fatality rates (CFRs) by age group



CFR(China) < CFR(Italy) !

Proportion of confirmed cases by age group

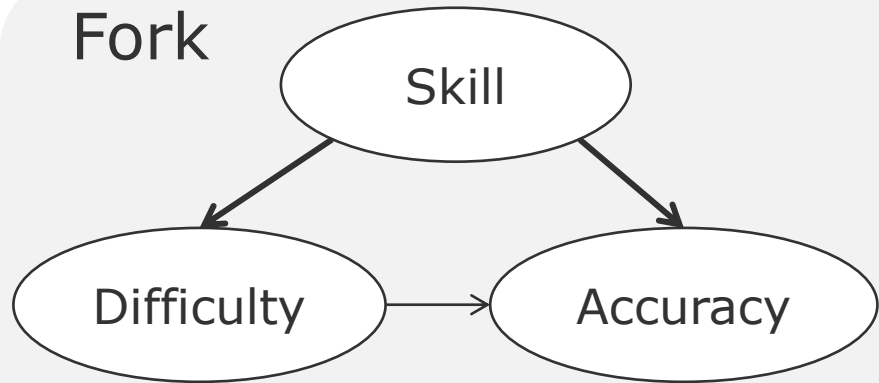


Because Italy has a larger proportion of confirmed cases among older age groups.

Solution – condition on additional factors = Age

When to condition? Confounding graph patterns

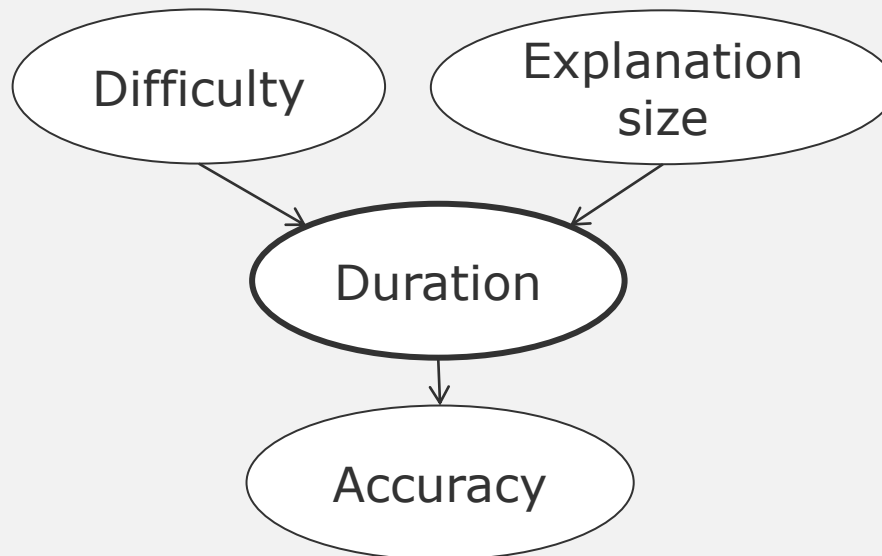
Fork



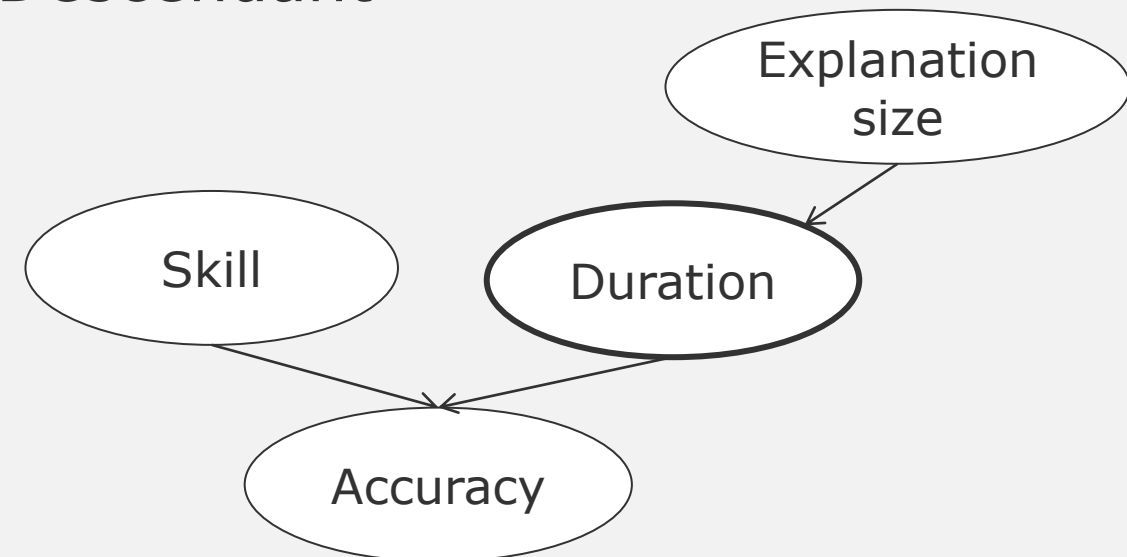
Pipe



Collider



Descendant



Reichenbach's Common Cause Principle

Assume that $X \not\perp\!\!\!\perp Y$ (X and Y are dependent)

- X causes Y
- Y causes X
- There is a third hidden common cause*
- Combination all the above

Hans
Reichenbach
1891-1953



In other words, "there is not correlation without causation"


*Also called **spurious correlation** between X and Y. Spurious correlations are not stable properties that hold across environments [Woodward 2005], hence invariance can be used to discover the causal structure [Arjovsky et al. 2019].

[Woodward 2005] Making things happen: A theory of causal explanation. Oxford university press.
[Arjovsky et al. 2019] Invariant risk minimization. *arXiv preprint arXiv:1907.02893*.

Condition on the parents (close backdoor paths)


Adjustment formula

outcome, e.g., True Positive


$$P(Y = y \mid do(X = x)) = \sum_z P(Y = y \mid X = x, Z = z) P(Z = z)$$



difficulty low (0) or high (1)



parents of X, e.g.,
programmer skill,
complexity of code,
buggy/not buggy

$$P(Y = y \mid do(X = x)) = \sum_z \frac{P(Y = y, X = x, Z = z)}{P(X = x, Z = z)}$$



Propensity Score

We looked at a single relationship, but we have multiple conditional dependencies.

The joint probability distribution for all n variable is given by:

$$P(x_1, \dots, x_n) = \prod_j P(x_j | x_1, \dots, x_{j-1})$$

Storing all probabilities requires a table with d^n or 2^n for binomial variables

Hence, the exhaustive exploration of all DAGs with n variables is super-exponential in n , hindering the effectivity of brute-force methods for observational causal discovery even for moderate n .

Brute force search does not scale

Nodes	Number of Distinct Directed Acyclic Graphs	Digits
2	3	2
3	25	2
4	543	3
5	29281	5
6	3781503	7
7	1138779265	10
8	783702329343	12
9	1213442454842881	16
10	4175098976430598143	19
11	31603459396418917607425	23
12	521939651343829405020504063	27
13	18676600744432035186664816926721	32
14	1439428141044398334941790719839535103	37
15	237725265553410354992180218286376719253505	42
16	83756670773733320287699303047996412235223138303	47
17	62707921196923889899446452602494921906963551482675201	53
18	99421195322159515895228914592354524516555026878588305014783	59
19	332771901227107591736177573311261125883583076258421902583546773505	66
20	2344880451051088988152559855229099188899081192234291298795803236068491263	73

7 nodes = 4 million graphs
8 nodes = 800 billion graphs

9 nodes = 1,548 times more
(or $1.2 \cdot 10^{16}$ graphs)

source: [Sloane 2019]

- Factorization as a Generative Model

- $P(x_1, \dots, x_{i-1}) = \prod_i P(x_i | \text{parents}(x_i))$

- where the set of $\text{parents}(x_j)$ are the Markovian Parents of x_j

- $P(x_1, \dots, x_{i-1})$ representation goes from exponential in **n** to linear in **n**

- $O(n2^k)$, k = maximum number of parents in the graph

However, this still requires two strong assumptions:

- Known structure of the graph (parent-child relationships) and
- Small maximum number of parents of any given node.

“The causal generative process of a system’s variables is composed of autonomous modules that do not inform or influence each other.

In the probabilistic case, this means that the conditional distribution of each variable given its causes (i.e., its mechanism) does not inform or influence the other mechanisms.”

Schölkopf, B. (2019). Causality for machine learning. arXiv preprint arXiv:1911.10500.

1. Knowledge can be decomposed in modules and mechanisms that are independent
2. Mechanisms can only be discovered by interventions, i.e., acting in the world
3. Mechanisms can be reuse across subdomains
4. Interventions usually affect only one mechanism at a time
5. There is only one causal graph that reflects the true causal mechanisms

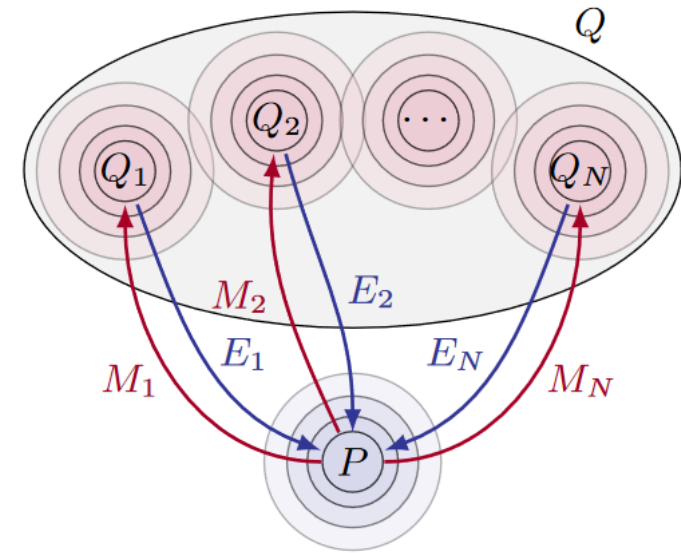
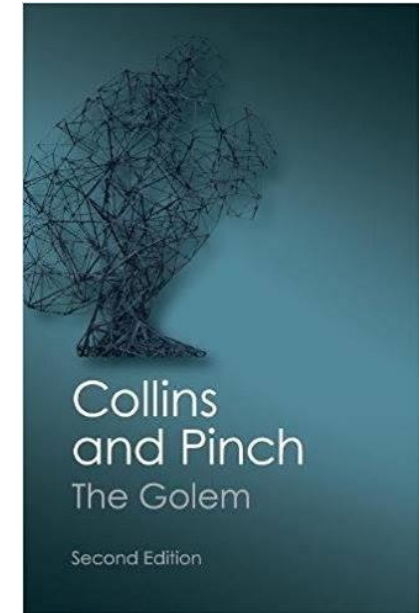


Figure 1. An overview of the problem setup. Given a sample from a canonical distribution P , and one from a mixture of transformed distributions Q_i obtained by mechanisms M_i on P , we want to learn inverse mechanisms E_i as independent modules. Modules (or *experts*) compete amongst each other for data points, encouraging specialization.

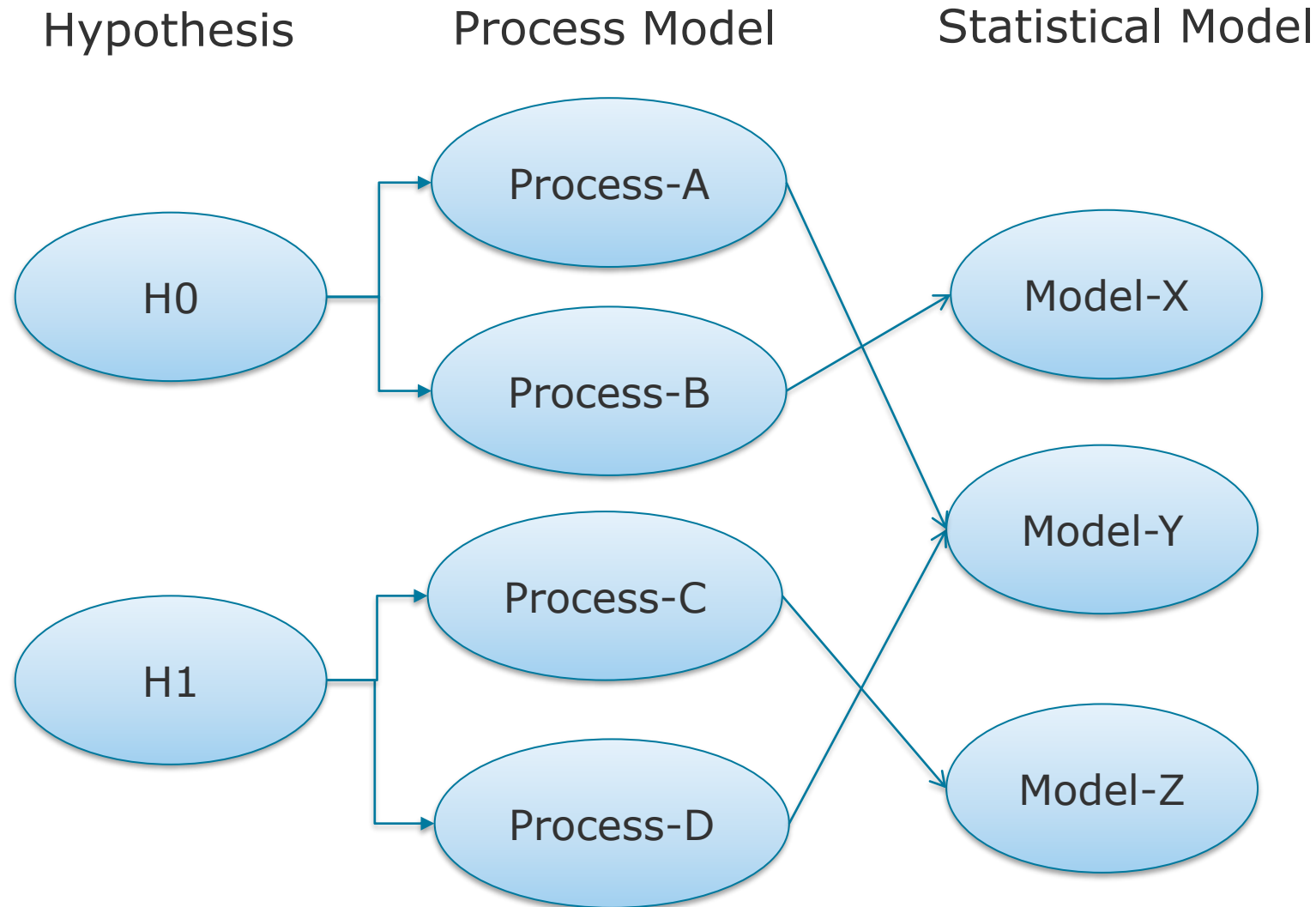
Statistical Inference

Statistical Modeling- Fundamental challenge

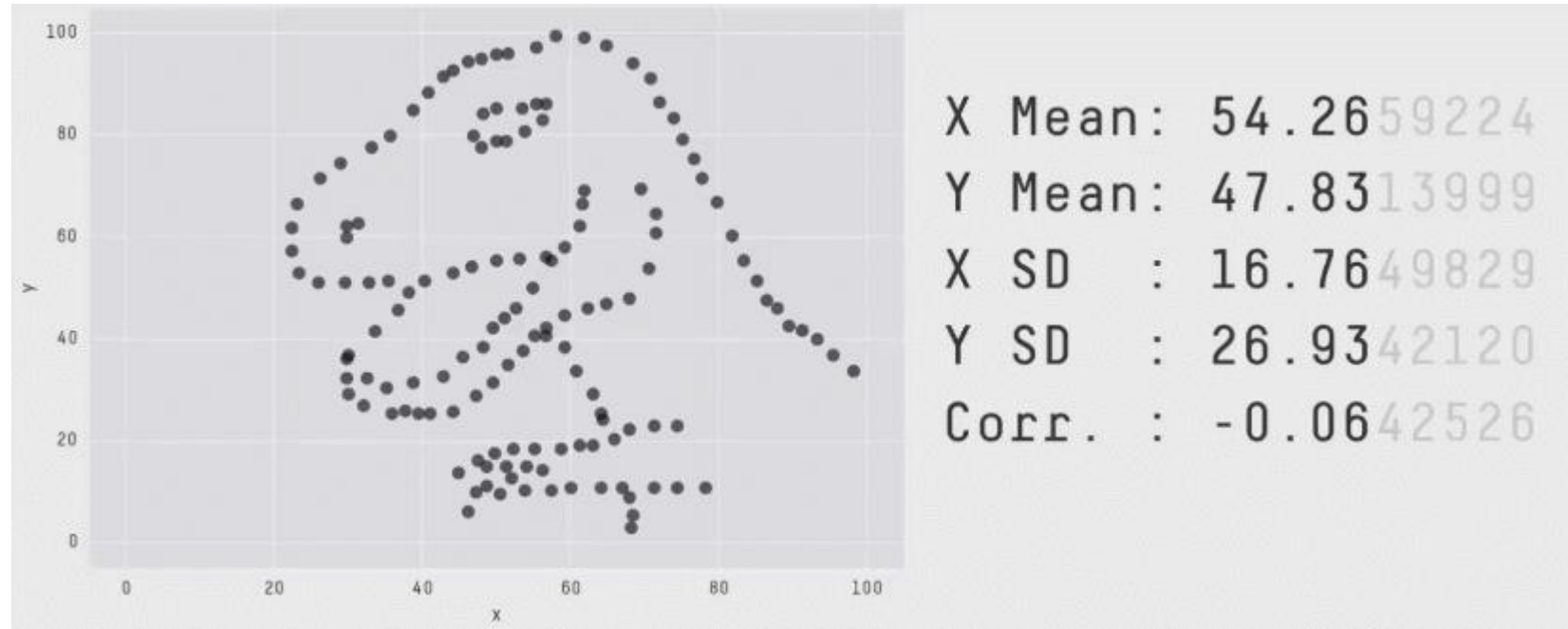
- Math & Engineering
- The Golem of Prague
- Characteristics
 - Animated by “truth”
 - Powerful
 - Ignorant to engineer's “intent”
 - Easy to misuse
 - Fictional / Always false



Statistical Modeling- hypotheses and models



Statistical Modeling- different processes same model



<https://www.autodeskresearch.com/publications/samestats>

- Hypothesis must be falsifiable
 - Process or model that can negate the hypothesis
- However
 - Hypothesis are not models
 - Measurement error
 - Latent variables
 - modus tollens
 - Denying the consequence
 - “If A then B, if **not** B then **not** A”

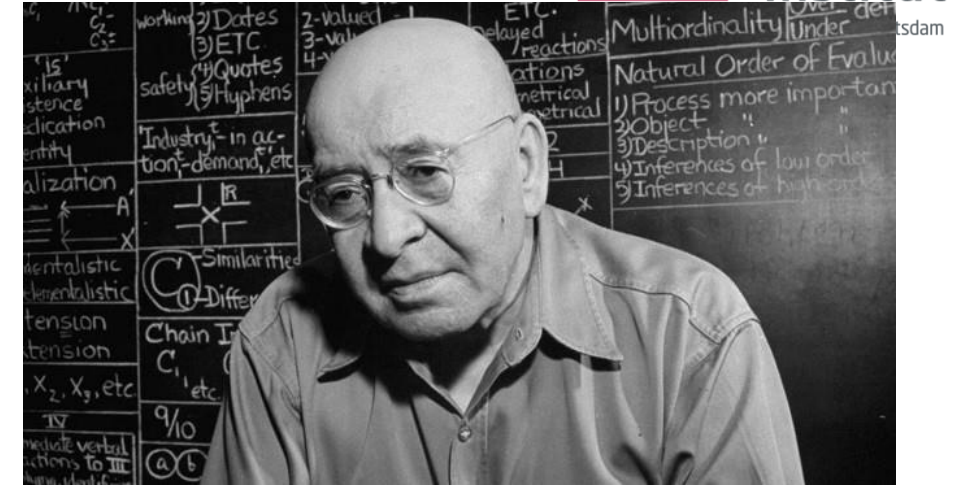
What we cannot know because it does not exist – Sean Carroll

- Ockham Razor
 - Highest rate of predictions by assumptions
 - Fewest assumptions for the same predictions

- Metrics
 - Error
 - Variance
 - Correlation
 - Area under the curve
 - Risk of Overfitting
 - Out of Distribution (OOD) Error
 - WAIC, Pareto Smooth Cross-Validation

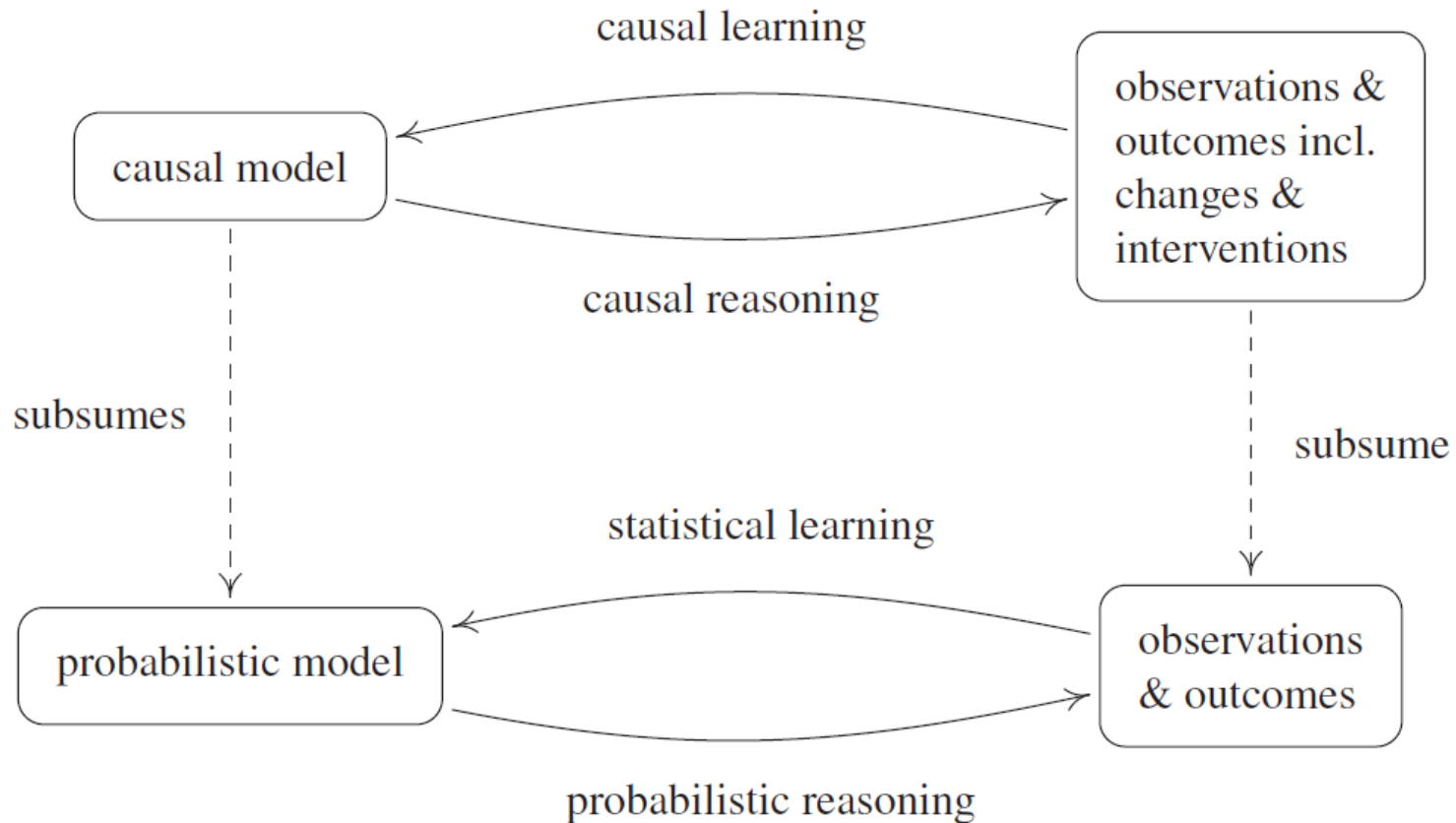
“A map is not the territory it represents, but, if correct, it has a similar structure to the territory, which accounts for its usefulness.”

— Alfred Korzybski, Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics



Towards Causal Mechanisms

Overall approach to causal inference

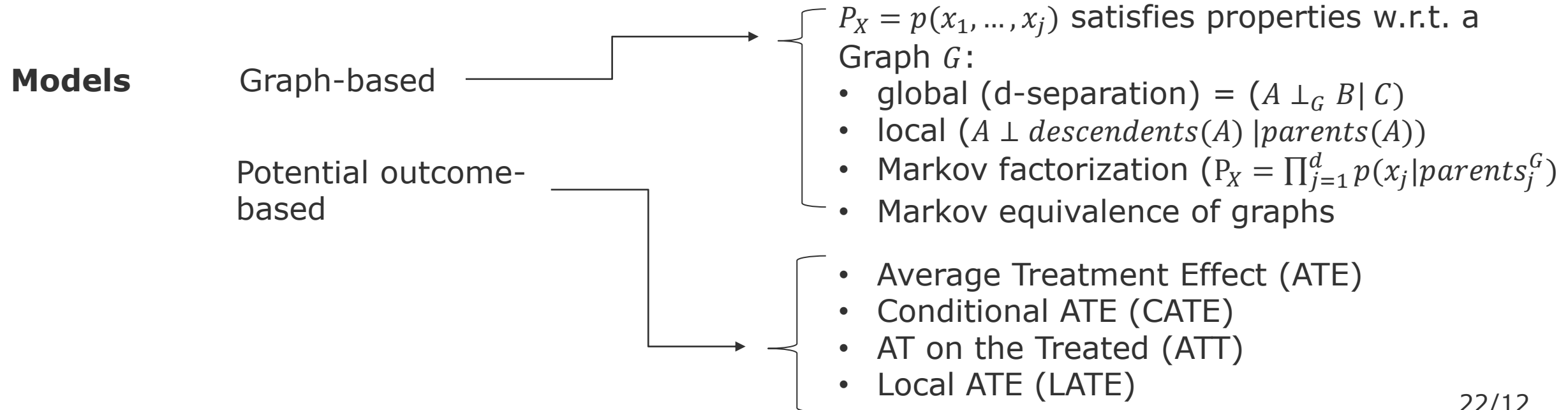


source: Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press.

Motivation	Fundamental problem of causality Confounding Reichenbach's principle
Assumptions	Ignorability (no confounders) Positivity (no imbalances) Modularity (stable unit of treatment) Consistency (compliance) Exclusiveness (effective intervention)

Methods

- Potential outcome
- No causation without intervention
- Causal discovery
- Matching
- Adjustments (backdoor criterion)
- Hypothetical interventions (do-operator)
- Instrumental Variables



Intuition: Interventions allow to maximize mutual information between:

- 1- intentions (goal-conditioned policies) AND
- 2- changes in state (trajectories) condition on the current state

Measures of dependence of mechanisms

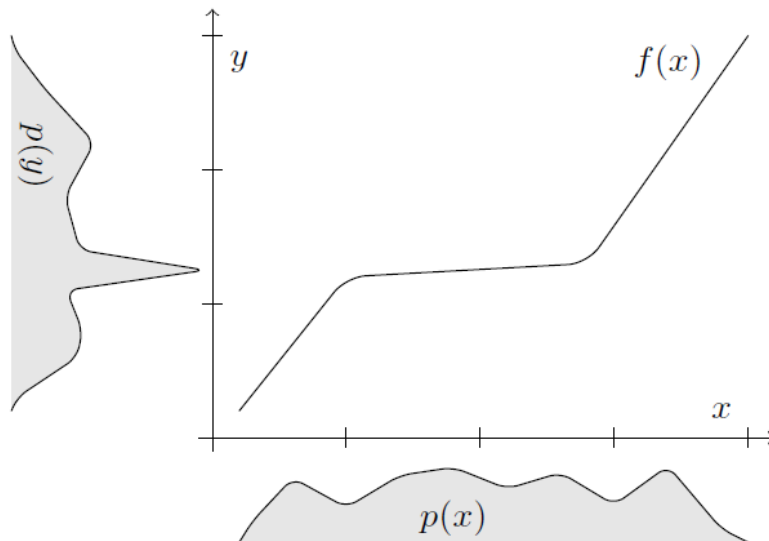


Figure 2: (from [Peters et al. \(2017\)](#)) If f and p_x are chosen independently, then peaks of p_Y tend to occur in regions where f has small slope and f^{-1} has large slope. Thus p_Y contains information about f^{-1} .

What should I do in the system to change Y?

What treatment should I take to attain a certain level of change on Y?

e.g., increase programmer skill to improve code inspection accuracy?

- Should recruit graduate student with minimum four years of experience OR
- Give programmers the double of the time to finish a task?
- Ask programmers to give a second or third look at the same source code?

What would have been the outcome Y if the covariate X had a different value?

e.g., would this programmer have been more accurate if she:

- Has spent more time?
- Had see the source code for a second and third time?
 - Hence, has had the chance to increase confident?
- Has written a longer explanation?

Definition:

modularity of the interventions consists of two assumptions:

1. the intervention on node X affects only the structural assignments of X
2. how the intervention is executed does not influence the results of the intervention

pg. 125 in [Peters, Janzing, & Schölkopf 2017]

Intuition:

We can set and unset variable values without altering the structure of the functions that relate the variables with each other.

Definition: As we vary the parameters of a causal model (e.g., regression coefficients), no independences in the probability model can be destroyed.

[Pearl 2009]

Intuition:

Causal relationships are more stable than probabilistic relationships, because we assume that

- Causal relationships = objective physical constraints in our world (ontological)
- Probabilistic relationships = belief or knowledge about the world (epistemic)

Definition:

It is possible to identify or estimate an unbiased result from data that was originated from selection bias, if the following holds:

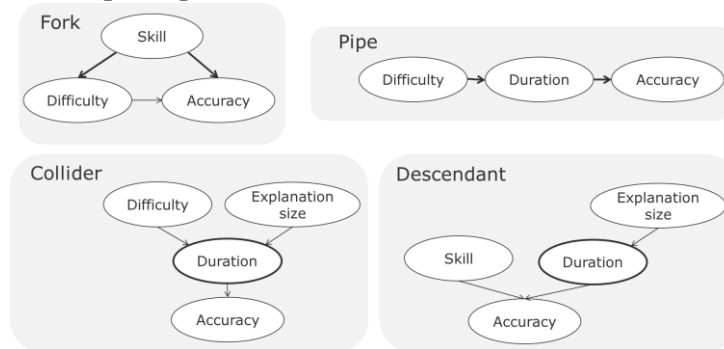
$Y(X)$ is conditionally independent of $X|Z$

Intuition:

Conditioning on Z takes away the variation of the individual. This means that the treated ($A=1$) and the control ($A=0$) are equivalent in terms the distribution of their covariates.

Overall approach to causal inference methods

Graph patterns



Markov factorization

$$P(x_1, \dots, x_n) = \prod_j P(x_j | x_1, \dots, x_{j-1})$$

Markovian Parents

$$P(x_1, \dots, x_{i-1}) = \prod_i P(x_i | \text{parents}(x_i))$$

$O(n2^{\text{parents}})$ instead of 2^n

Conditional Expectation

$$E(X|y) = \sum_x xP(x|y)$$

Regression

$$y = b + ax$$

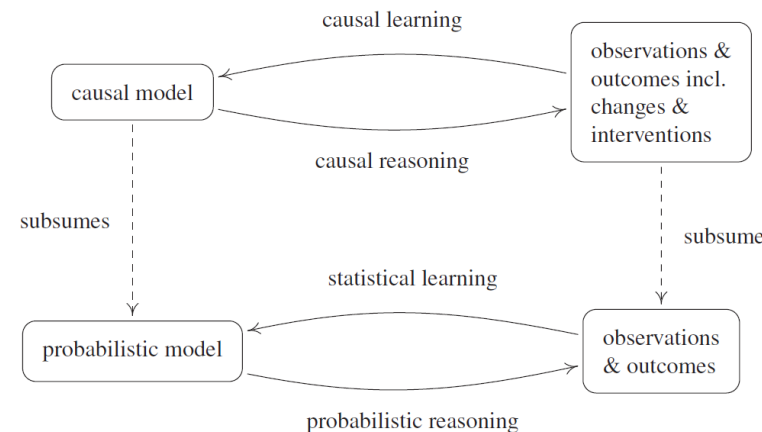
$$\alpha = R_{YX|Z} = \rho_{XY|Z} \frac{\sigma_{X|Z}}{\sigma_{Y|Z}}$$

Do-operator and Intervention

$$P(Y = y | do(X = x)) = \sum_z P(Y = y | X = x, Z = z) P(Z = z)$$

Potential Outcome

$$P(Y = y | do(X = 1)) - P(Y = y | do(X = 0))$$



Causal Inference Assumptions:

- Ignorability (no confounders)
- Positivity (no imbalances)
- Modularity (stable unit of treatment)
- Consistency (compliance)
- Exclusiveness (effective intervention)

Structural Equation Model

$$x_1 = u_1$$

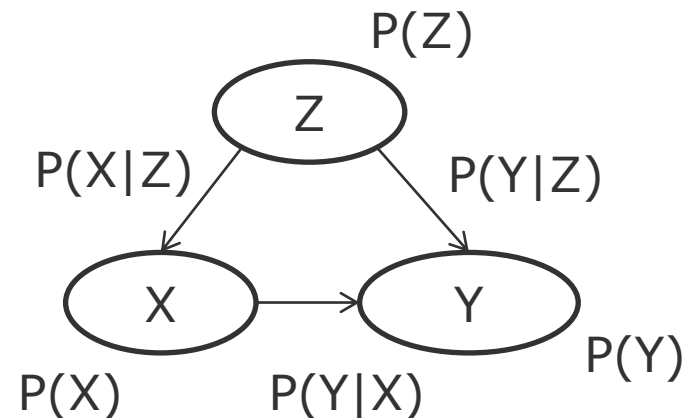
$$x_2 = \alpha_1 x_1 + u_2$$

$$x_3 = \alpha_{3,1} x_1 + \alpha_{3,2} x_2 + \alpha_{3,3} x_1 x_2 + u_3$$

Fundamental Concepts

Definition: A Bayesian Network (BN) is a Directed Acyclic Graph (DAG) where nodes are propositional variables and edges are local dependencies between conceptual related propositions.

Intuition: Nodes are variables and edges are conditional probabilities.



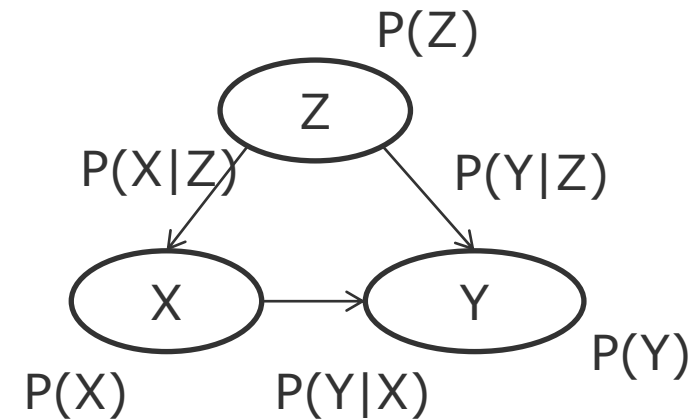
Recap - Conditional probability and Bayes Theorem

$$(1) \quad P(Y|X) = \frac{P(Y, X)}{P(X)}$$

$$(2) \quad P(X|Y) = \frac{P(X, Y)}{P(Y)}$$

$$(3) \quad P(Y|X) = \frac{P(X|Y) P(Y)}{P(X)} \quad (\text{Bayes Theorem})$$

$$(4) \quad P(Y|X) = \frac{P(X|Y) P(Y)}{\sum_x P(X|Y) P(Y)}$$



(Mean) $E(X) = \sum_x xP(x)$

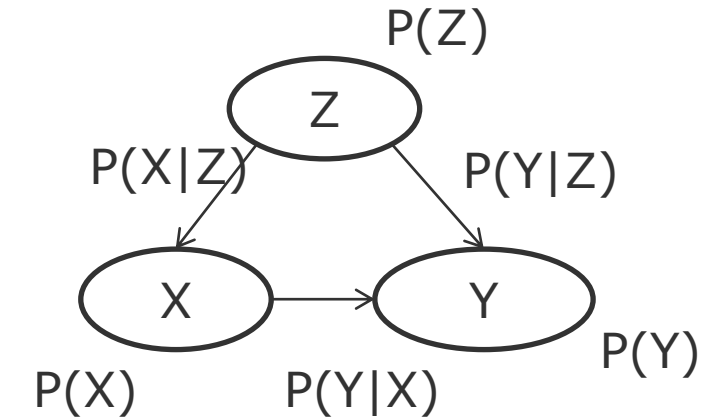
(Conditional Mean) $E(X|y) = \sum_x xP(x|y)$

(Expectation of a function) $E[g(Y, X)] = \sum_{x,y} g(x, y)P(x, y)$

(Covariance) $g(Y, X) = \sigma_{XY} = E[(X - E(X)) \cdot (Y - E(Y))]$

(Correlation) $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

(Conditional Correlation) $\rho_{XY|Z} = \frac{\sigma_{XY|Z}}{\sigma_{X|Z} \sigma_{Y|Z}}$



(Regression)

$$y = a + bx$$

$$b = R_{YX|Z} = \rho_{XY|Z} \frac{\sigma_{X|Z}}{\sigma_{Y|Z}}$$

$$E[Y | do(X = x)] \neq E[Y | X]$$

Because of the unobserved confounders!

Can I predict the outcome of an intervention?

A causal quantity is identifiable if it can be written as a function of the observed variables.

In other words. If I observe an infinite amount of data from my processes can I make a prediction under intervention?

Can I use the observable joint distribution to make a prediction about some other joint distribution (which is the one that I intervened)?

- Markov Factorization
- Reichenbach's Common Cause Principle

Scalability &
Stability

- d-Separation
- Markov-Blanket
- Causal Markov Condition

Modularity

- do-Operator
- Back-door criteria
- Adjustment

Ignorability

$$P(x_1, \dots, x_n) = \prod_i P(x_i | \text{Parents}(x_i))$$

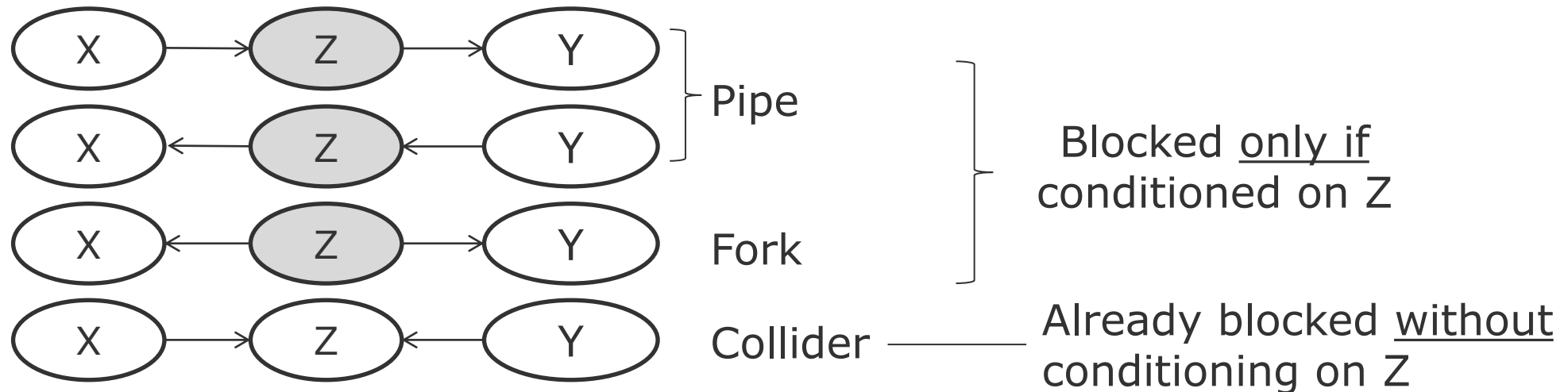
 Causal Markov kernels

If P admits the factorization relative to a DAG G , we can say that the DAG represents the probability function P , i.e., G and P are compatible or that P is Markov relative to G .

We can also say that the Graph G induces the probability P .

X and Y are d-separated if all paths between X and Y are blocked by a set Z of nodes,
i.e., $X \perp Y \mid Z$

The blocking* situations are:



*grey means blocked (usually by conditioning on it)

Consequences of d-Separation

Definition:

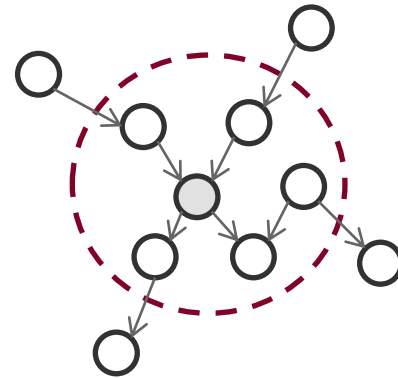
- P is Markov w.r.t. G for all Z , if X is d-separated of Y given Z , i.e., $X \perp Y \mid Z$ [Peters, Janzing & Schölkopf 2017]
- Hence, for all distinct variables X and Y in G , if X does not cause Y , then
$$P(X \mid Y, \text{Parents}(X)) = P(X \mid \text{Parents}(X))$$
 [Hausman & Woodward 1999]

Intuition:

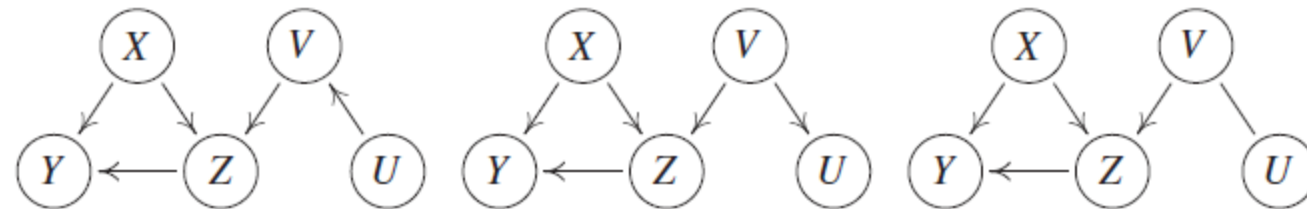
Conditional on its parents, X is independent of every variable in G except its effects [Spirtes, Glymour & Scheines 1993]

Definition: A Markov Blanket of X consists of all $\text{Parents}(X)$, $\text{Descendants}(X)$, and $\text{Co-Parents}(X)$

Intuition: A node is conditionally independent of the entire network, given its Markov blanket.



- Graphs are equivalent if they have the same set of nodes and present the same colliders.



source: [Peters, Janzing & Schölkopf 2017]

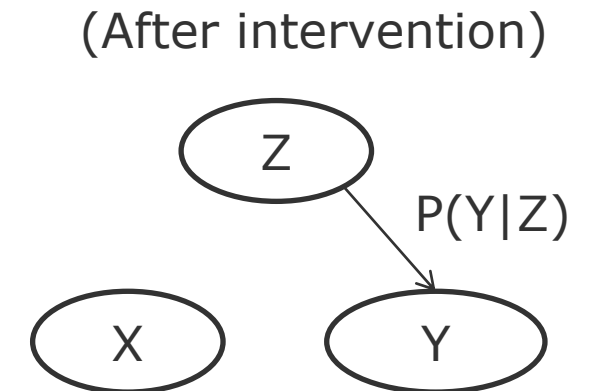
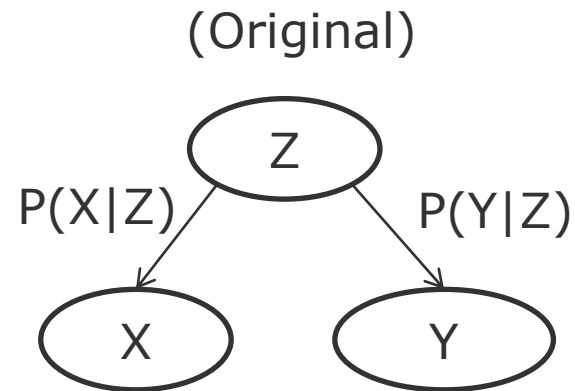
Definition:

$P(y|z, x)$ = What is the distribution of Y given that I observe $X = x$, $Z=z$?

$P(y|z, do(x))$ = What is the distribution of Y if I could set the value of X to x but let Z vary according with the original process?

Intuition:

$P(y|z, x) \neq P(y|z, do(x))$



Definition: A set of variables Z satisfies the back-door criterion relative to X, Y if no node in Z is a descendant of X , and Z blocks every path between X and Y that contains an arrow into X .

Then the causal effect of X on Y is given by:

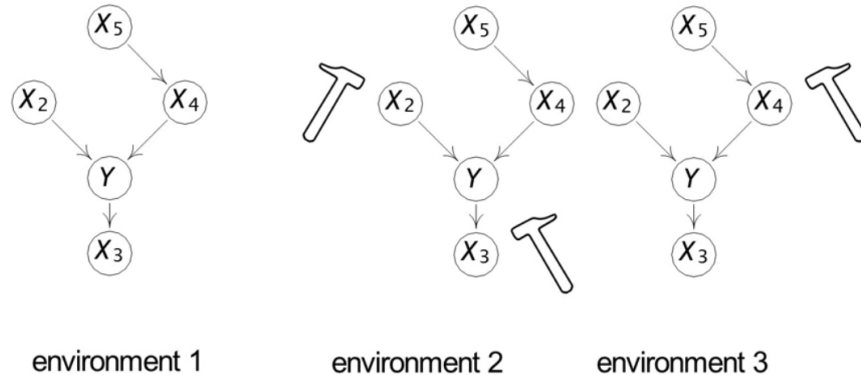
$$P(Y = y \mid do(X = x)) = \sum_z P(Y = y \mid X = x, Z = z) P(Z = z)$$

Intuition:

1. We want to block back-door paths because they make X and Y dependent.
2. However, we do not want to block any descendants of X because they might mediate causal effect from X to Y.
3. We do not want to create spurious paths, e.g., by conditioning on a collider that creates a spurious dependency between X and Y.
4. Since these dependencies are not causal, they are confounders.
5. Back-door criterion enables to attain the ignorability condition, hence mitigate selection bias [Morgan & Winship 2015].

Principal of Invariant Prediction [Peters, Bühlmann & Meinshausen 2016]

If one knows the values of the causes of an outcome of interest, then it is irrelevant how these values came about.

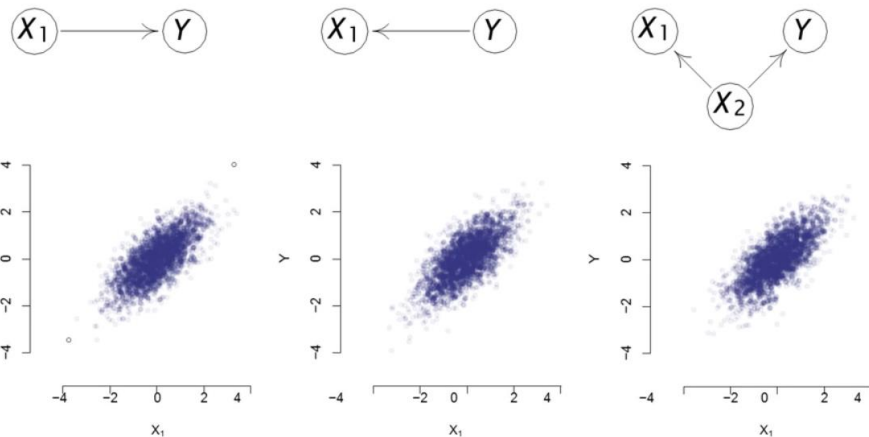


Invariance of $Y | X_S = x$ holds if we take $S = S^*$ as complete set of direct parents (here $S^* = \{X_2, X_4\}$).

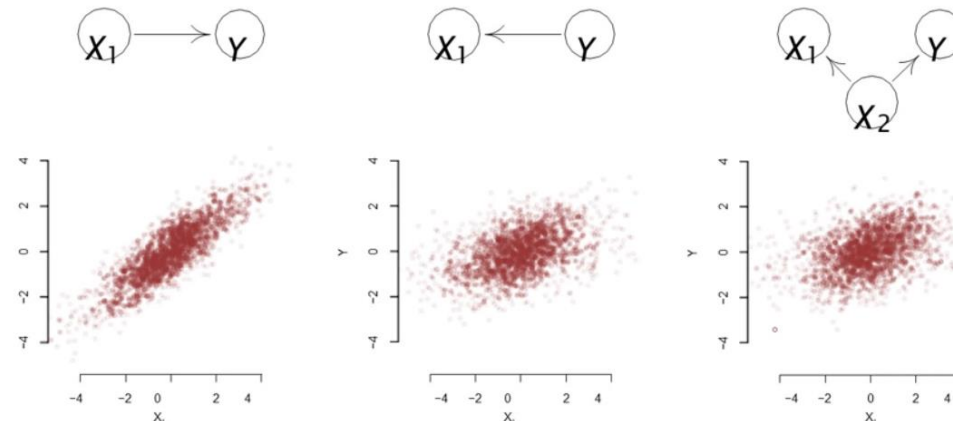
Modularity, Stability, and Consistency

- Causal inference requires that assumptions about invariance to operationalize interventions
- While we do not have agree on the set of interventions, we need to theoretically support them.
- Interventions can be of at least two types:
 - noise intervention (perturb a variable) or
 - do-interventions (set a variable to a certain value).

Different causal models same observational distribution



Distribution for observational data and under noise-intervention on X1



Causal Discovery with Neural Networks: in presence of latent variables, confounding, and sparsity

Most machine learning work essentially tries to skip the opening steps, tackling complex problems empirically, without ever trying to build a firm understanding about what initial primitives are really required for language and higher-level cognition.

Skipping those first steps has not gotten us thus far to language understanding and reliable trustworthy systems that can cope with the unexpected; it is time to reconsider.

Gary Markus (2020), **The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence**

Why do we need Causality in Machine Learning?

Context: Correlation-based models make four strong assumptions

1. Out Of Distribution (OOD)
 - Covariate Shift and Concept Drift (non-stationarity)
2. Independent and Identically Distributed (IID) data
 - Scrambling Training and Testing data
3. Lack of guidance to intervention (control and exploration)
4. Credit assignment (explanation, counterfactual thinking)
 - Search through possible futures is highly intractable

Impact:

- Limits systematic generalization
- High complexity for adaptation

1. Understand how distributions change (simulate non-stationary and dynamic processes)
2. Composable local representations
 - Deep nets layers
 - Factorization of Joint Probabilities
 - Sparse factor graphs
3. Alternative assumptions (causal mechanisms) to the I.I.D. assumption
 - Independence of mechanism principle
 - Identify knowledge (rules) that can be reused
 - meta-learning and transfer learning
4. Temporal Causal Mechanisms
 - Explanations only for short sequences

1. Learning Functional Causal Models with Generative Neural Networks [Goudet et al. 2018]
2. A meta-transfer objective for learning to disentangle causal mechanisms [Bengio, Y., et al. 2019]
3. Recurrent Independent Mechanisms [Goyal et al. 2020]
4. Learning Neural Causal Models from Unknown Interventions [Ke et al. 2020]

My Reading Wishlist on Causality in Neural Nets

Out of Distribution Learning

Zhang, Kun, et al. "Domain adaptation as a problem of inference on graphical models." arXiv preprint arXiv:2002.03278 (2020).

Gong, Mingming, et al. "Causal generative domain adaptation networks." arXiv preprint arXiv:1804.04333 (2018).

Ton, Jean-Francois, Dino Sejdinovic, and Kenji Fukumizu. "Meta learning for causal direction." arXiv preprint arXiv:2007.02809 (2020).

Yuan, Ye, Xueying Ding, and Ziv Bar-Joseph. "Causal inference using deep neural networks." arXiv preprint arXiv:2011.12508 (2020).

Scalable Models

Kriváchy, Tamás, et al. "A neural network oracle for quantum nonlocality problems in networks." npj Quantum Information 6.1 (2020): 1-7.

Anker, Justin J., et al. "Causal network modeling of the determinants of drinking behavior in comorbid alcohol use and anxiety disorder." Alcoholism: clinical and experimental research 43.1 (2019): 91-97.

Wang, Yuhao, et al. "Causal discovery from incomplete data: a deep learning approach." arXiv preprint arXiv:2001.05343 (2020).

Wu, Pengzhou, and Kenji Fukumizu. "Causal mosaic: Cause-effect inference via nonlinear ica and ensemble method." International Conference on Artificial Intelligence and Statistics. PMLR, 2020.

Pawlowski, Nick, Daniel C. Castro, and Ben Glocker. "Deep structural causal models for tractable counterfactual inference." arXiv preprint arXiv:2006.06485 (2020).

Zhu, Rong, et al. "Efficient and Scalable Structure Learning for Bayesian Networks: Algorithms and Applications." arXiv preprint arXiv:2012.03540 (2020).

Zhang, Yulai, et al. "Parallel ensemble methods for causal direction inference." Journal of Parallel and Distributed Computing (2021).

Dynamical Models

Xie, Xiao, Moqi He, and Yingcai Wu. "CausalFlow: Visual Analytics of Causality in Event Sequences." arXiv preprint arXiv:2008.11899 (2020).

Kurthen, Maximilian, and Torsten Enßlin. "A Bayesian Model for Bivariate Causal Inference." Entropy 22.1 (2020): 46.

Garrido, Sergio, et al. "Estimating Causal Effects with the Neural Autoregressive Density Estimator." arXiv preprint arXiv:2008.07283 (2020).

Wu, Hang, Wang. "An Information Theoretic Learning for Causal Direction Identification." 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). IEEE, 2020.

Inferring Causality in Feedback Systems by Combining Neural Nets and Ordinary Differential Equations [Schoenberg 2020]

Epistemology of Causal Deep Learning

Vallverdu, Approximate and Situated Causality in Deep Learning, (2020)

Zenil, et al., Causal deconvolution by algorithmic generative models. Nat. Mach. Intell (2019)

Zenil, et al., An Algorithmic Information Calculus for Causal Discovery and Reprogramming Systems, iScience. (2019)

Parafita & Vitrià , Causal Inference with Deep Causal Graphs (2020)

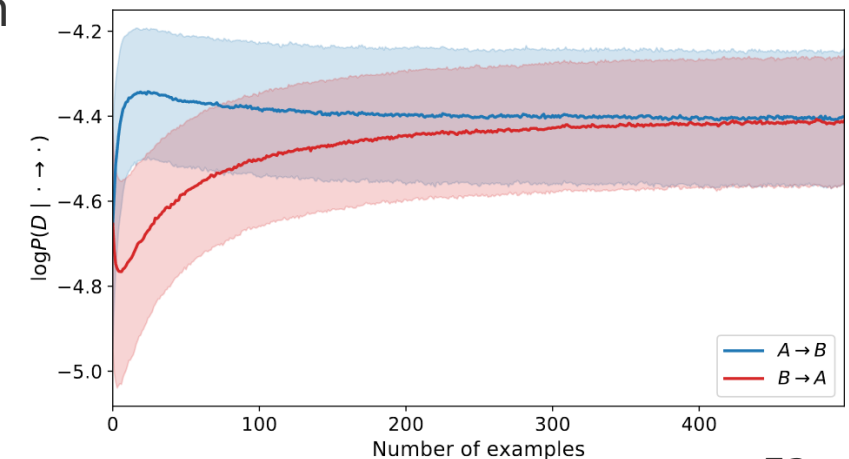
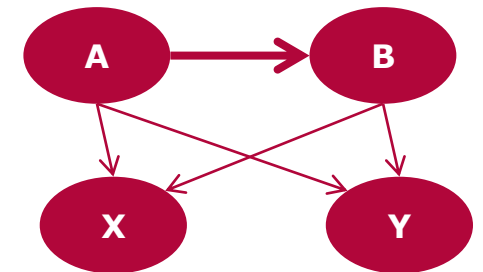
Causal Generative Neural Networks [Goudet et al. 2017]

CGNNs are causal models of the world that are able to simulate the outcome of interventions to discover v-structures and causal-effect relationships.

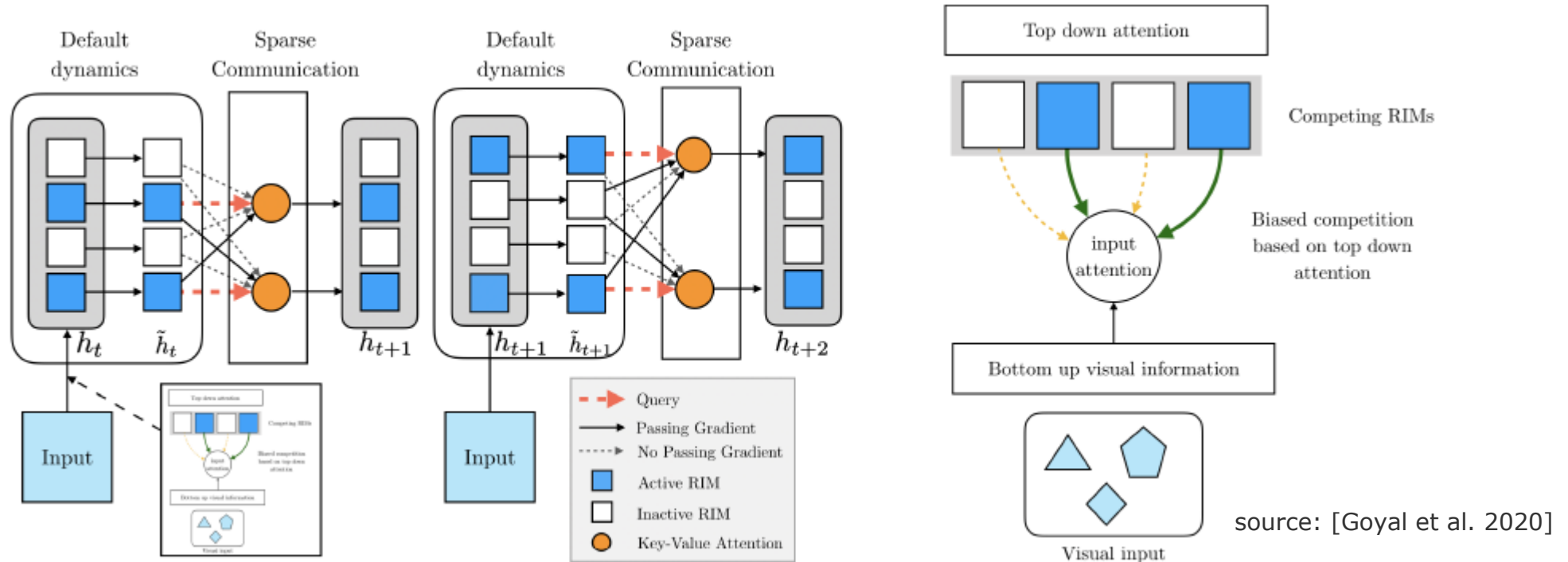
Meta-Transfer Objective for Learning to Disentangle Causal Mechanisms [Bengio et al. 2019]

Factorizes a Joint Distribution by using OOD. The learner that has the right factorization will adapt much faster (i.e., require less data). Adapt consists to recover to a change in the distribution

- A causes B or vice versa? How to disentangle (A,B) from observed (X,Y)
- Approach: Exploit changes in the distribution AND Speed of adaption



Goal: learn modular structures that (1) reflect the dynamics of the environment and (2) better generalizes, i.e., is more robustness to changes on few of the underlying causes. Approach: Modularize the computation.



first stage, individual RIMs produce a query which is used to read from the current input

second stage, an attention-based competition mechanism is used to select which RIMs to activate (right figure) based on encoded visual input (blue RIMs are active, based on an attention score, white RIMs remain inactive)

third stage, individual activated RIMs follow their own default transition dynamics while non-activated RIMs remain unchanged

fourth stage, the RIMs sparsely communicate information between themselves, also using key-value attention.

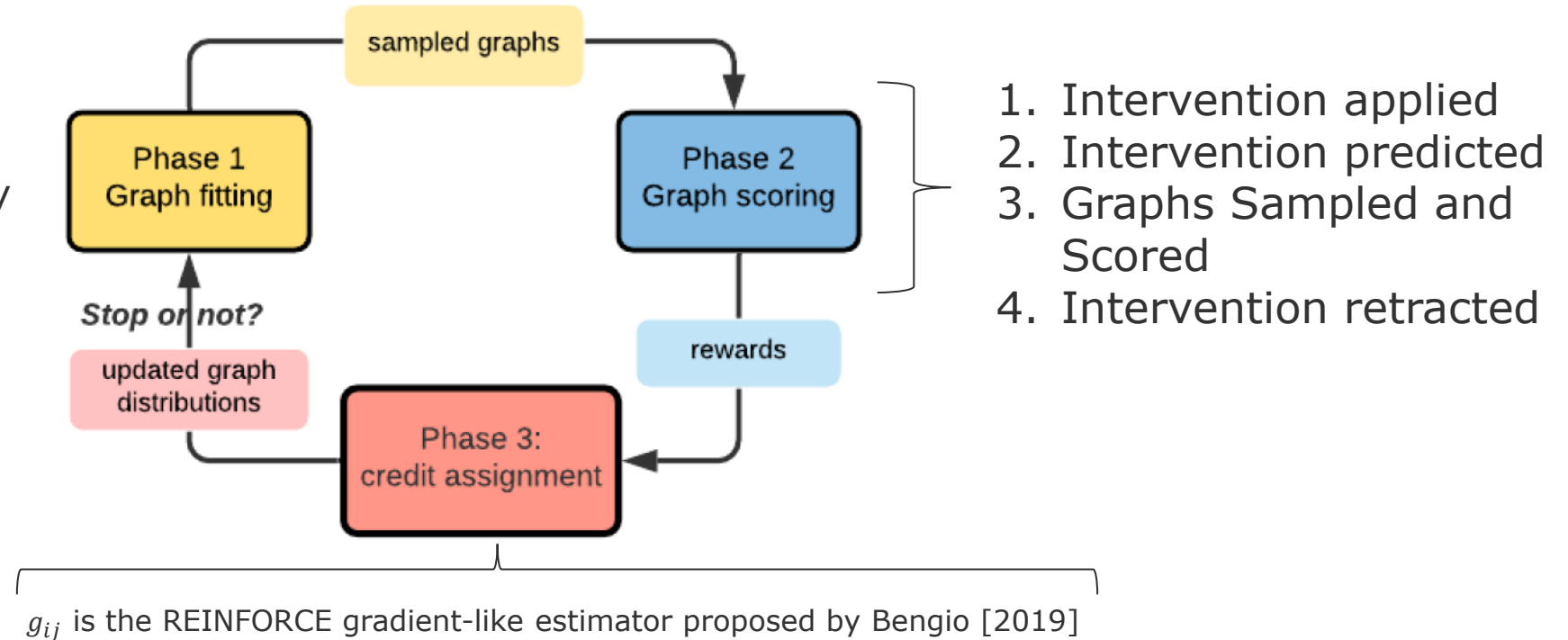
Works better than traditional methods

Learning small causal graphs helps avoid exponential

Functional parameters are trained to maximize the likelihood of randomly drawn observational data under graphs randomly drawn from our current beliefs about the edge structure.

$$C_{ij} \sim \text{Ber}(\sigma(\gamma_{ij}))$$

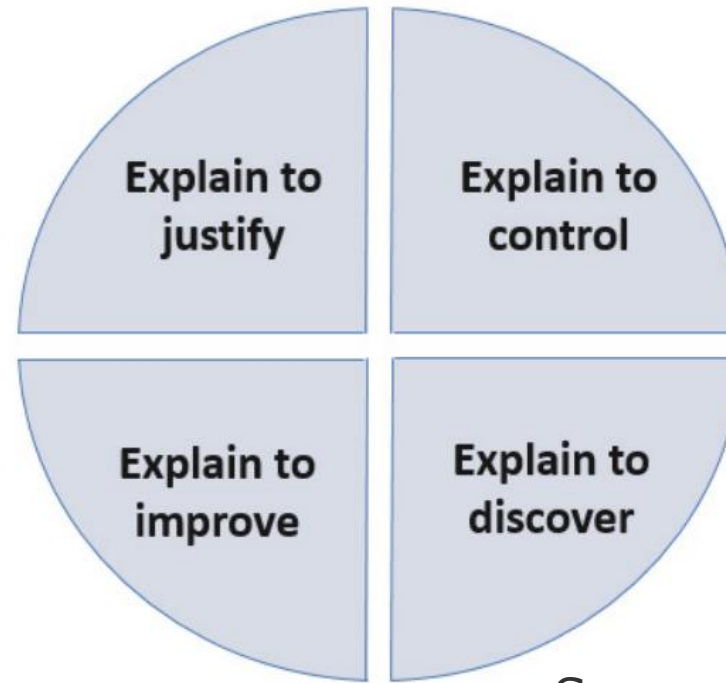
hypothesized configuration of the SCM's DAG



g_{ij} is the REINFORCE gradient-like estimator proposed by Bengio [2019]

$$g_{ij} = \frac{\sum_k (\sigma(\gamma_{ij}) - c_{ij}^{(k)}) \mathcal{L}_{C,i}^{(k)}(X)}{\sum_k \mathcal{L}_{C,i}^{(k)}(X)}, \quad \forall i, j \in \{0, \dots, M-1\}$$

$\mathcal{L}_{C,i}^{(k)}(X)$ is the Log-likelihood of log-likelihood of variable X_i in the data sample X under the k 'th configuration, $C^{(k)}$, drawn from our edge beliefs.



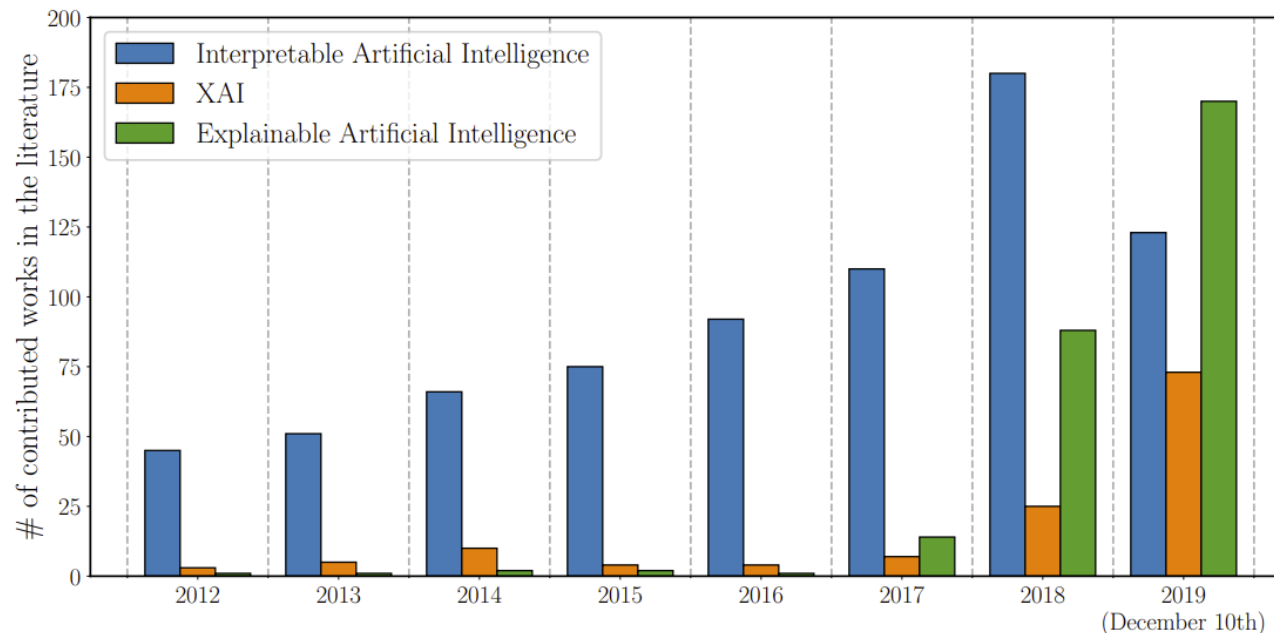
Source: [Adadi & Berrada 2018]

Causal Inference for Explainable Neural Networks:
Trust, Fairness, Transfer Learning

Causality for Explainable AI

eXplainable AI (XAI) are a set of ML techniques to [Arrieta et al 2020]

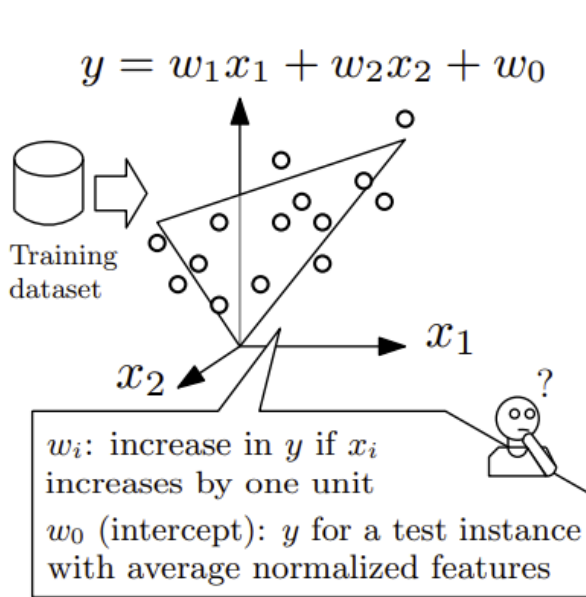
- produce more explainable models while maintaining a high level of learning accuracy
- enable humans to understand, trust, and effectively manage the new artificially intelligent partners



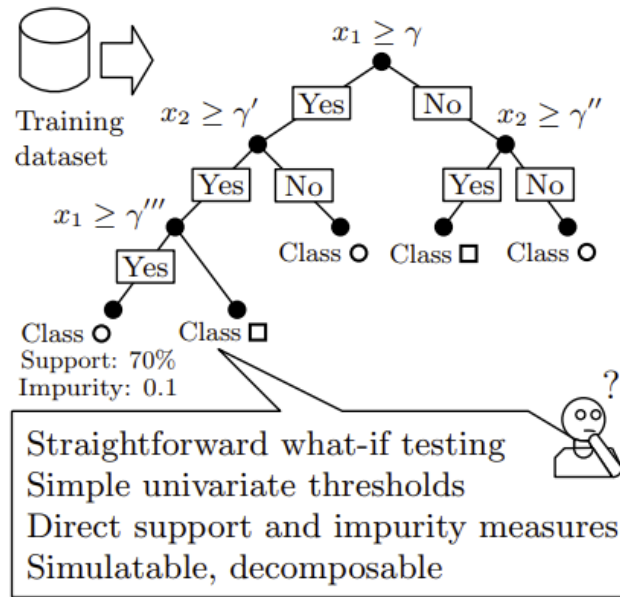
source: [Arrieta et al. 2020]

XAI Goal	Main target audience (Fig. 2)
Trustworthiness	Domain experts, users of the model affected by decisions
Causality	Domain experts, managers and executive board members, regulatory entities/agencies
Transferability	Domain experts, data scientists
Informativeness	All
Confidence	Domain experts, developers, managers, regulatory entities/agencies
Fairness	Users affected by model decisions, regulatory entities/agencies
Accessibility	Product owners, managers, users affected by model decisions
Interactivity	Domain experts, users affected by model decisions
Privacy awareness	Users affected by model decisions, regulatory entities/agencies

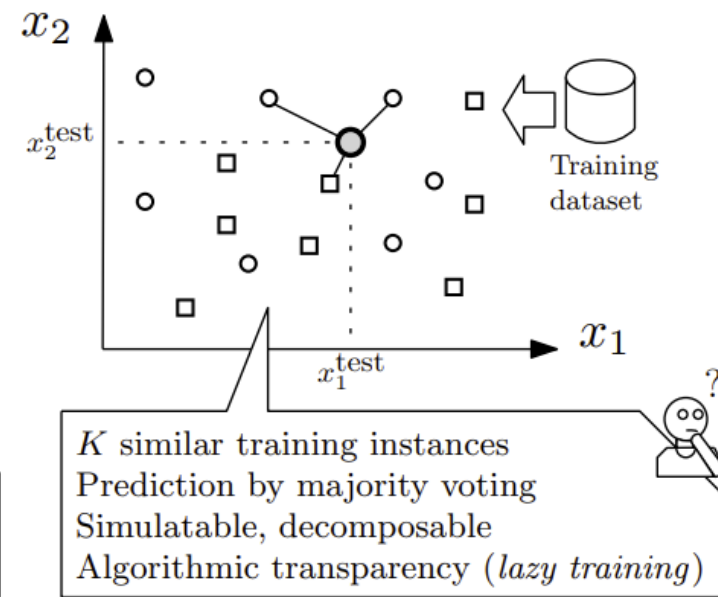
Levels of transparency of ML models [Arrieta et al. 2020]



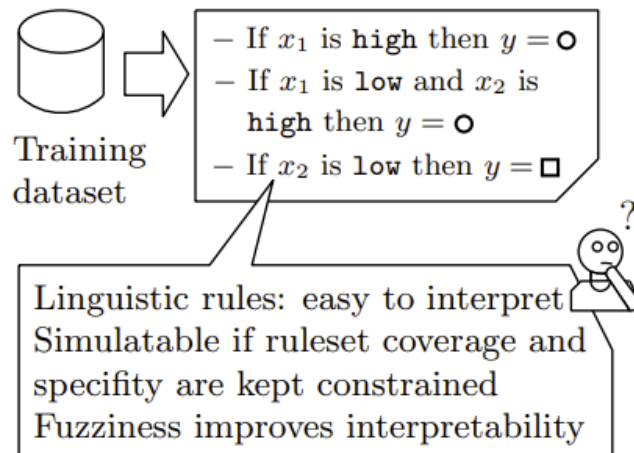
Linear regression



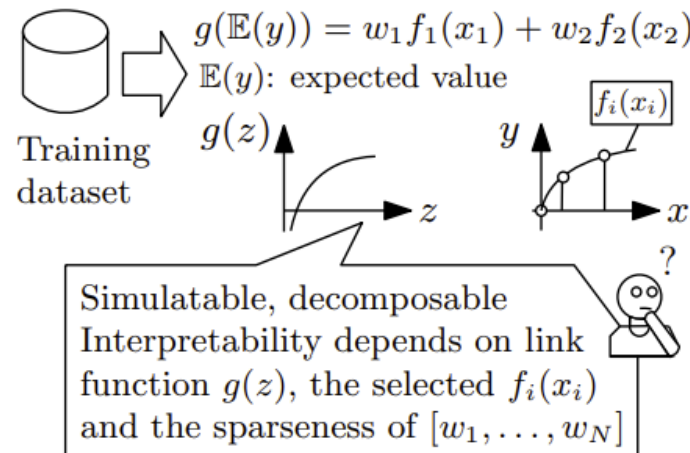
Decision Tree



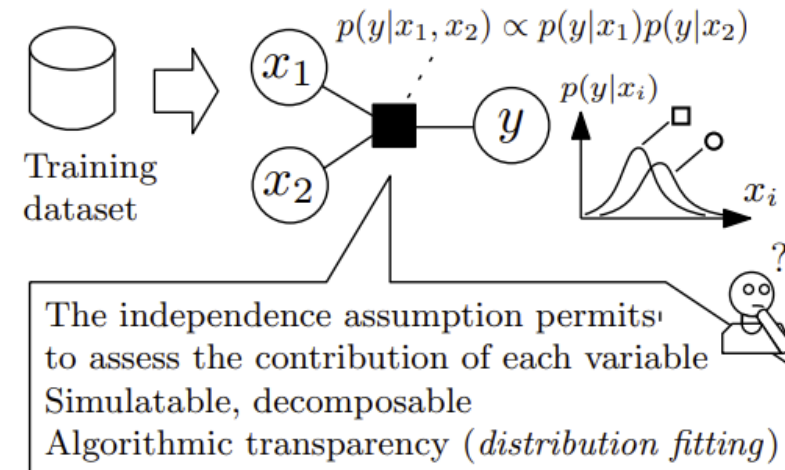
K-Nearest Neighbor



Rule-Based Learners

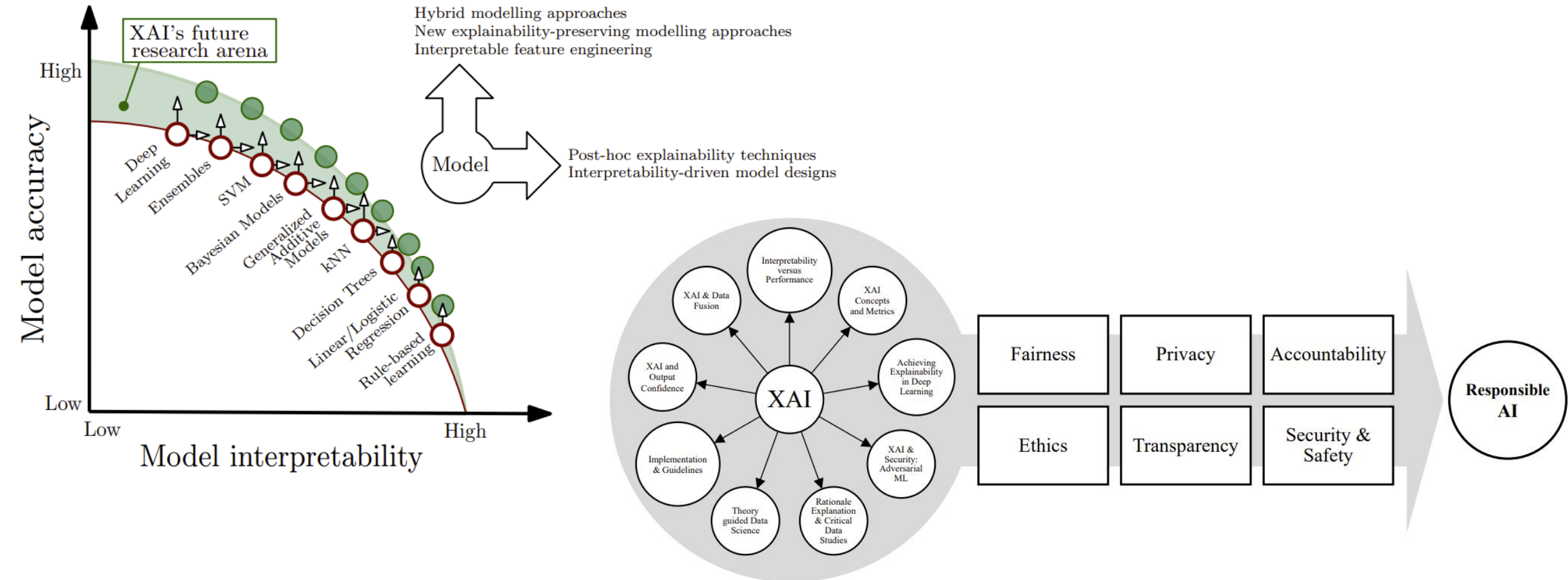


Generalized Additive Methods

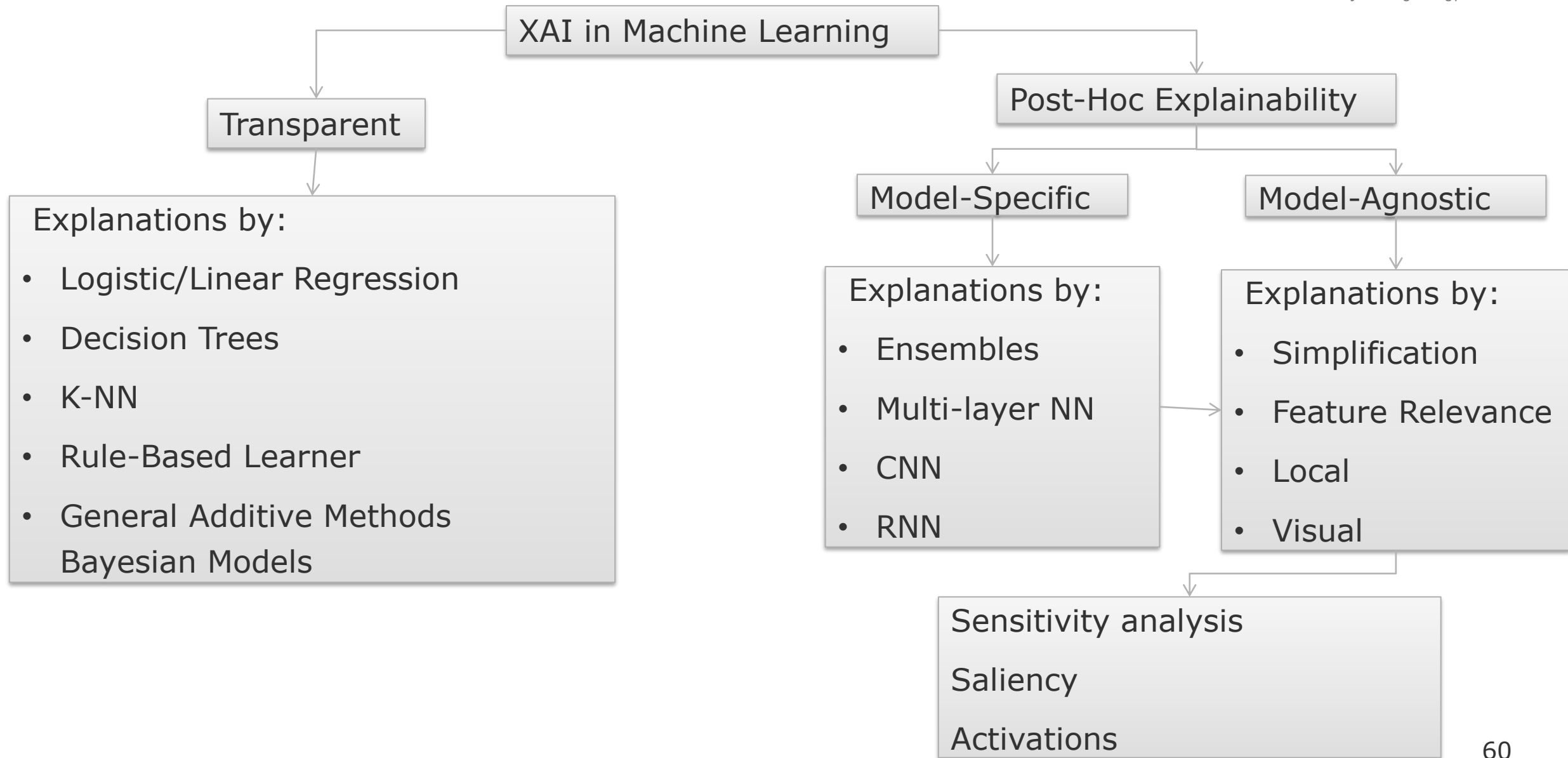


Bayesian Models

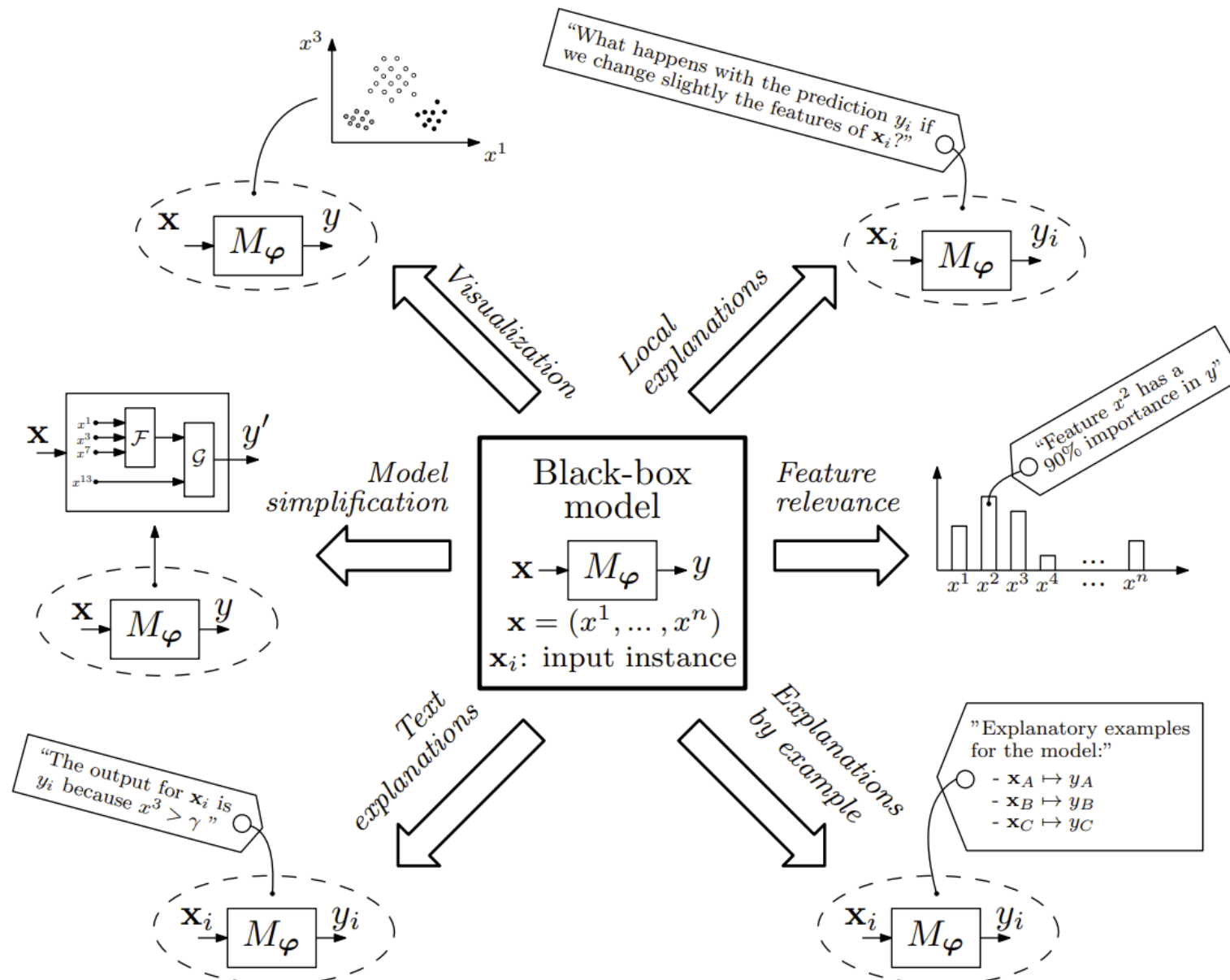
XAI Deep Learning and Summary [Arrieta et al. 2020]



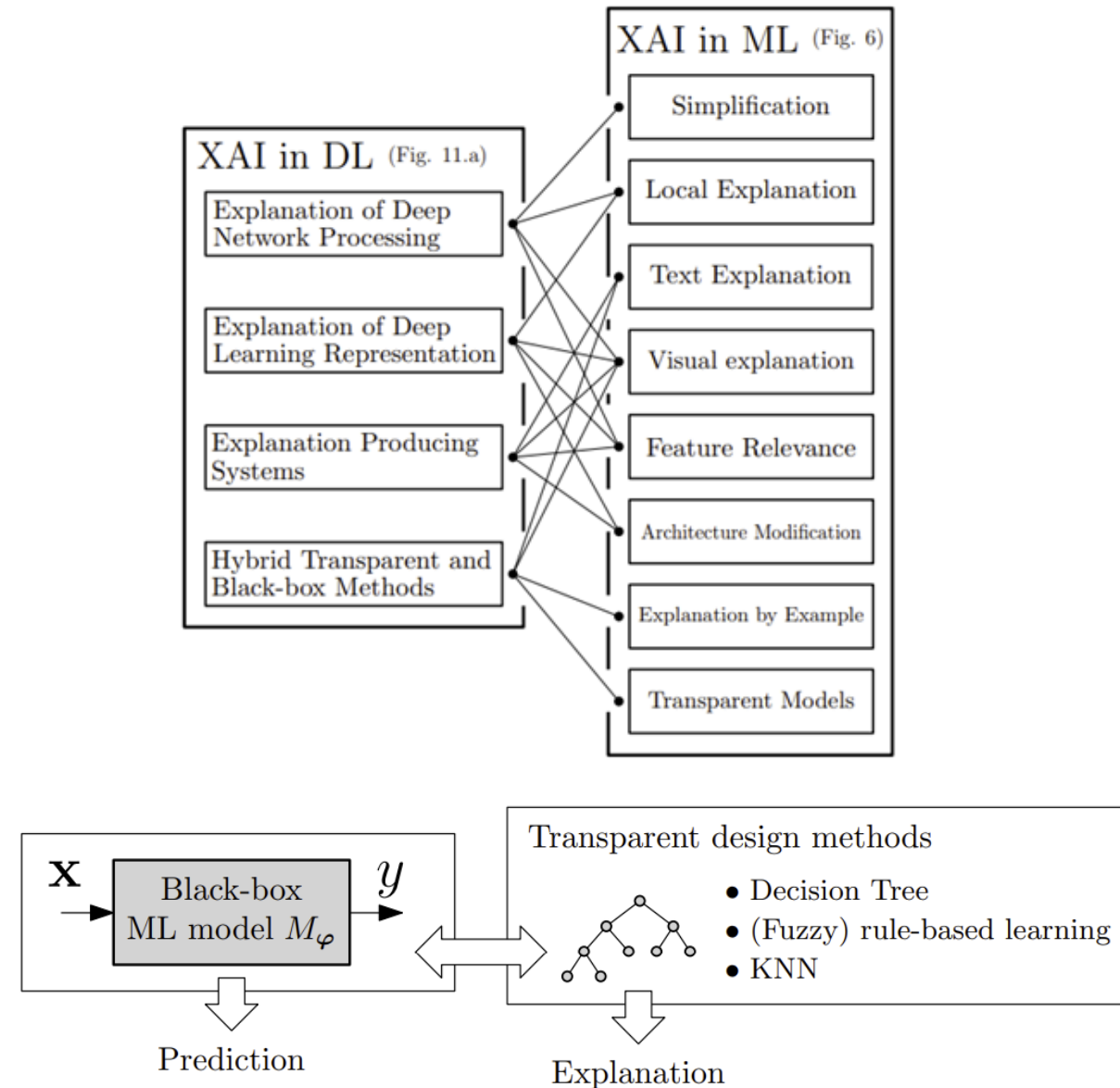
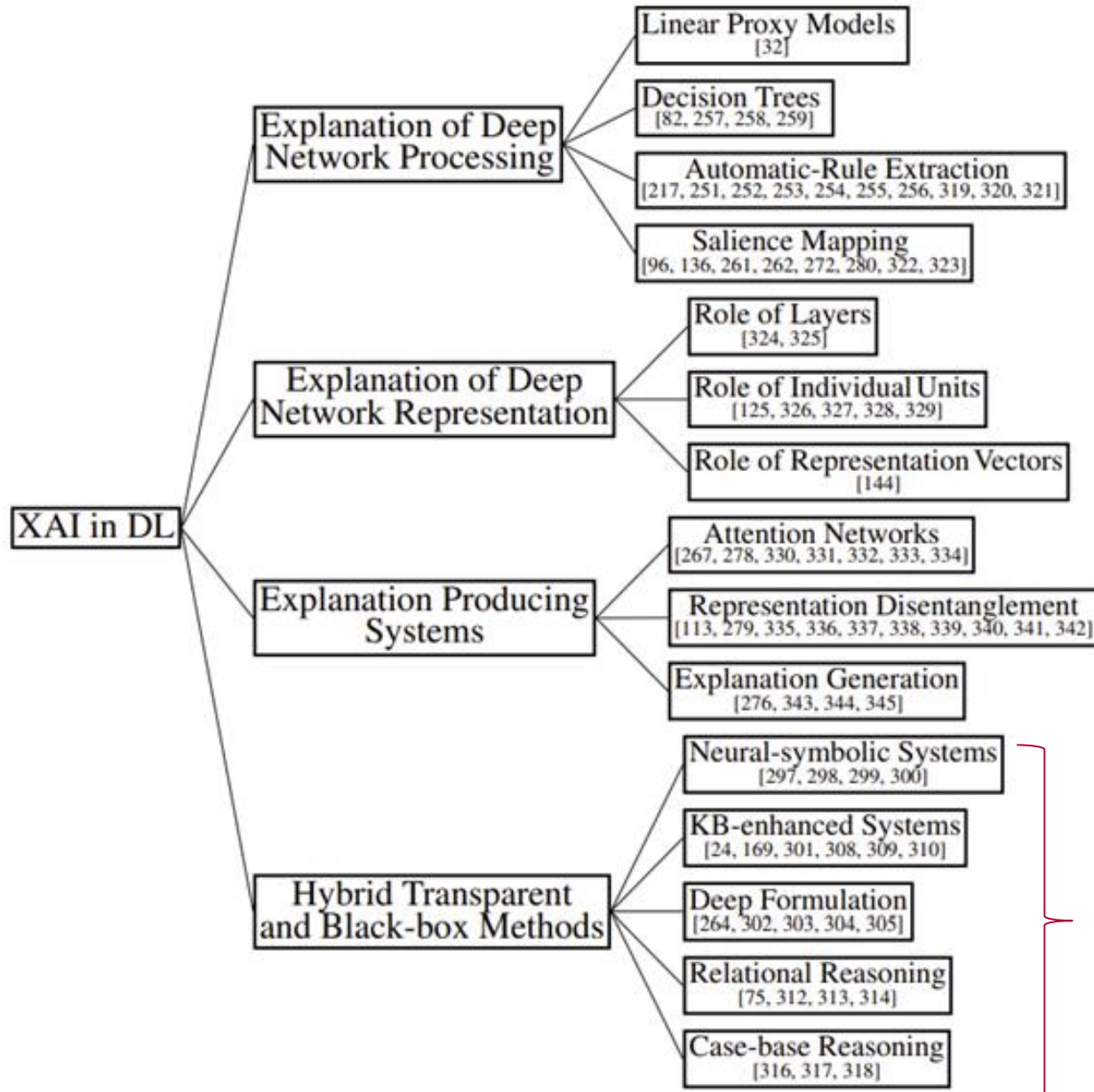
source: [Arrieta et al. 2020] Arrieta, A. B., et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information Fusion* 58 (2020): 82-115.



Pos-hoc Explainability [Arrieta et al. 2020]



XAI in Deep Learning [Arrieta et al. 2020]



Reflection: Beyond accuracy vs explainability trade-off

- GNN allowed graphs to be input of Neural Networks
 - This extended the graph representation from computation to feature engineering
- Attention Mechanisms allowed to use the input graph to adapt the computation graph
 - This extended the memory that a neural network can hold
- Causal inference could extend GNN and DL to:
 - Use graph computation to continuously discover causal structure
 - Out of Distribution, Inductive learning
 - Adapt the computation graph online
 - Invariance learning, Principled Auto-ML
 - Develop a more principled design approach to deep learning
 - Pipeline Abstractions, Design Patterns [1,2,3], Frameworks

Key Take-Away Insights

1. Causality is invariant
2. Causality allows to exploit correlations safely
3. Causal structure mitigates for sparsity and supports transfer learning

[1] Zenil, et al., 2019, Causal deconvolution by algorithmic generative models. Nat. Mach. Intell

[2] Zenil, et al., 2019, An Algorithmic Information Calculus for Causal Discovery and Reprogramming Systems, iScience.

[3] Gustafsson, Vallverdu, 2015, The Best Model of a Cat Is Several Cats. Trends Biotechnol.

1. Please contact me to schedule your final presentations
2. Continue individual meetings for project update
 - Level of detail of outcomes to report and discuss
 - Format of the presentations
 - Content of the final paper
3. Final Lecture on Propagation Graph Neural Networks March 17, 5pm.

END