A Review on Approaches for Causal Structure Identification

Josephine Rehak

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
Josephine.Rehak@kit.edu
ORCiD: 0000-0001-6139-9703

Abstract

Learning the skill to discover causal relations and to make use of them is said to be an essential step in human intelligence [43, 31] and potentially also in machine intelligence [28, 38]. The domain of causal discovery tackles the challenge of identifying causal structures from data collected from observations or experiments by exploiting special properties of causal relations. While current causality literature focuses on methods of probabilistic discovery using conditional independence tests and hard and soft interventions[27, 5, 19] other lesser known approaches are neglected [33].
In this work, we will give a short review on approaches for gaining causal knowledge and provide a categorization of methods. Also, we will introduce the Joint Discovery Assumption that is essential for combining different approaches for causal discovery. Finally, we discuss the open research fields we deduce from our categorization.

1 Introduction

Would it not be great if artificial intelligence (AI) could understand the cause and effects of its past actions and adapt accordingly? What sounds like a dream may be supported by causal-understanding AI in the future. What is called *causal discovery* or *causal structure learning* may level the path for causal reasoning in AI. The strength of these methods lies in finding causal relationships, while also being able to distinguish correlation from causation which is a weakness of many current machine learning methods[38]. For us humans, we are easily able to understand and use causal connections since infancy[22]. The step to causal understanding was a major breakthrough in human evolution [43, 31] the same may be true for the evolution of AI [28, 38].

Currently, such tasks are still beyond the possibilities of an AI as current algorithms fail to keep up with human capabilities. They struggle in simple tasks as when trying to discover the causal connection between the altitude of a weather station above sea level and its average outdoor temperature [15]. The actual causal connection is clear for us humans to see, because we know the average temperature cannot influence the altitude of a weather station.
Our assumption is that future algorithms will have to combine several existing approaches for discovery to gain the most causal structure information. In this paper, we will give a review on such approaches to gain what we call causal structure clues: information about the presence or absence of causal relations in a causal graph, by considering research in various science domains. We contribute by:

- categorizing algorithms to gain causal structure clues.
- providing a common ground for joint inference over various science domains by the definition of the Joint Discovery Assumption.
- deriving prospects for future research.

We cover methods used by humans, but also more advanced methods based on probabilities and spacetime which commonly challenge human understanding. We do not investigate methods for identifying whole causal networks from the

given structure clues as this would exceed the scope of this work.
In section 2, we give a short overview on the fundamentals on general causal structure learning. In section 3, we introduce a fundamental assumptions for the joint use of multiple approaches for causal discovery. The categorization itself is explained in section 4. We provide a short analysis of future prospects based on made categorization in section 5 and summarize our findings in section 6.

2 Basics of Causal Structure Discovery

In causal discovery, one tries to fully discover the true causal graph G^* for a given process under investigation. Such a graph consists of a set of variables V depicted as nodes and a set of edges E connecting the variables. It is a common assumption that G^* is always a causal directed acyclic graph (DAG); causal, as all edges in E indicate causal relations between the variables in V; directed, as all edges in E are only allowed to be only one-directed or absent; acyclic, as the edges in E may not form any cycles like bi-directed relations. Cycles may only occur in temporal considerations.
For given structure knowledge over the absence or direction of edges of G^*, a set of equally possible DAGs can be inferred which form an equivalence class of DAGs with respect to the provided knowledge. These equivalence classes can be represented by Partially Directed Acyclic Graphs (PDAGS). If no existing knowledge is present, the equivalence class contains all feasible DAGs for given variables. For this case, the number of DAGs is described in [41]. By adding structure knowledge to an equivalence class, one may direct or eliminate edges and thereby restrict the equivalence class. G^* is deemed to be fully discovered if all possible edges between the variables of V are discovered to be either absent or one-directed [27].
Approaches to gain such causal structure knowledge are explained in section 4 in detail.

3 Joint Discovery Assumption

We humans naturally use multiple approaches to identify causal relationships. This is based on the assumption that no matter which causal discovery method is used the underlying graph is the same as long as no undetected changes have occurred in the graph. As long as each structural clue is implied by the causal graph the discovered causal clues cannot contradict each other unless when essential assumptions of the discovery methods were disregarded.
As an example, if two causal discovery methods m_1, m_2 infer a causal relation between the variables A and B results in two structural clues c_1 and c_2, then c_1 must be supported by c_2 and c_2 must be supported by c_1 accordingly. This means the discoveries $c_1 = \{A \rightarrow B\}$ and $c_2 = \{A \perp\!\!\!\perp B\}$ are not be possible without one of the clues to be erroneous. But, for example $c_1 = \{A \rightarrow B\}$ and $c_2 = \{A \not\perp\!\!\!\perp B\}$ are two structural clues that fully support each other.
We call this Joint Discovery Assumption, since it allows the joint use of several discovery methods. Based on this assumption, we can combine various structure clues from data to gain the most information on the true causal graph and we can evaluate more different types of data, since results from matching discovery algorithms can be included in the discovery process. In addition, computationally inexpensive methods can be used to restrict computationally or cost intensive discovery methods. Such an application could lead to a considerable streamlining, since previously excluded causal relationships no longer need to be considered by the second method. Note however that not every causal discovery method could infer every structural clue, since each comes with its own presuppositions.

4 Approaches for Discovering Causal relations

In the following, we grouped methods by similarity in their approach of exploiting properties of causal relations for inferring causal structure clues.

4.1 Structure clues from expert knowledge

The easiest discovery approach for artificial intelligence to learn causal relations is to get them taught by humans. We define such a structure clue as any knowledge that can be used for the construction of causal graphs. This may include knowledge over specific known edges [24], a known causal order [9], a known partial causal order [36], a known path between variables [3] or even knowledge over variable types [4]. We will deal with the latter in particular.
Typing assumptions [4] can only be applied if the investigated variables contain variables of similar type, e.g. multiple diseases or multiple drugs for treatment, and if structure knowledge over the type in relation to other variables is given, e.g. variables of type 'disease' may cause variables of type 'symptoms'. Such a typification of variables has to be performed by domain experts in advance. The general background knowledge over types can be used to limit the causal structure search. For example the provided domain knowledge may entail that diseases may cause symptoms, but symptoms may not cause diseases. The causal discovery methods can then consider these general rules in their detailed search. This example requires manual typification of the variables though investigation in automated type classification of variables may deliver promising results [4]. Using expert knowledge in general is a comfortable way of combining prior domain knowledge with the power of discovery algorithms. It is also an easy way to speed up discovery by reducing the amount of edges that need to be considered in causal structure search while leaving the actual causal discovery task to an algorithm. Unfortunately, introducing expert knowledge comes with the risk of introducing structure faults which can result in a wrong causal graph.

4.2 Structure clues from probabilistic inference

Since Judea Pearl published the foundations for probabilistic causal discoveries in 1988, the domain of probabilistic discovery is flourishing. By collecting data via observations with additional knowledge over the variable states, we are able to recover probabilities. Depending on the type of algorithm, these are interpreted differently. In the domain, the categorization into score-based and constraint-based algorithms has become established [11]:

Constraint-based methods make use of patterns in the retrieved probabilities to uncover fragments of the causal structure. For one, unconditional independence tests allow the learning of skeletons by identifying the presence, but not the direction, of causal relations. Further on, conditional independence tests can uncover immoralities, sometimes also called uncovered colliders or v-structures, by making use of d-separation properties [29]. Score-based algorithms try to find the graph with the highest fit to the data by varying edges in the graph. The accuracy of fit is measured in a likelihood score as the Bayesian information criterion [2] or the Aikaike information criterion [1]. A common example is the Greedy Equivalence Search (GES) [7].
The third category forms a group of mixture methods that use both score-based and constraint-based learning like Max-Min Hill Climbing (MMHC) [45].
All these approaches can be used to learn Markov Equivalence Classes (MEC), equivalence classes of DAGs which are equivalent in their conditional and unconditional independencies [46]. MECs can be represented in the form of complete partially directed acyclic graphs (CPDAGs) [42]. For each algorithm the resulting MECs may differ depending on found conditional independencies. Unfortunately, MECs can still include countless DAGs, especially when investigating big data [14]. Hence, the current challenge in this domain lies within finding the structure knowledge to reduce the MEC further. For this purpose, some publications resort to other discovery approaches to orient the remaining edges for example by using interventions [10] (see section 4.3) or noise methods [26] (see section 4.5). All in all, using probabilistic discovery we can discover important elements of the true causal graph, but may not discover it completely. This comes at the disadvantage that the discovery relies completely on the availability of huge quantities of unbiased data. Without such large sample sets, it is hard to gain profound results from (un)conditional independence tests [29].

4.3 Structure clues from manipulation

Another approach to discover causal relations that comes natural for humans is by experimenting with manipulations of variables. With a chosen intervention, a variables state or probability of occurrence is changed. This may trigger a change in the causally dependent variables which can be measured and

allows conclusions about causal dependencies. This builds on the fundamental assumption that an intervention targeted on the cause may influence the effect, but an intervention targeted at the effect may never create a change in the cause. An essential aide for such experiments are ceteris paribus conditions where the environmental variables can be recreated so several interventions may be tried out. These interventions are commonly of the following two types: Hard interventions, also called Pearlian interventions or structure interventions, forcefully set a variable to a chosen value and thereby eliminate all influences from other causes. They were first formalized in Judea Pearls do-calculus [30]. This calculus was proven to also fully support the popular potential outcomes framework [34].
So called soft interventions introduce an additional variable which causally affects the target variable and changes its probability of occurrence without disrupting the influence of other variables [5, 19]. For both kinds, the numbers of interventions required for the full identification of a true causal graph were identified by [10]. Also, several methods were established to identify the effects created by these interventions. Assuming strong ceteris paribus conditions, the effect of structure interventions for example can be discovered and measured by the Average Treatment Effect (ATE). It is calculated as the normalized sum over the individual treatment effects of all individuals or samples $i = 1, ..., N$. The individual treatment effect is the difference of the treated outcome variable $y_1(i)$ and the untreated individual outcome variable $y_0(i)$.

$$\text{ATE} = \frac{1}{N} \sum_i (y_1(i) - y_0(i))$$

In simple scenarios, few samples of each interventional outcome may suffice to identify the causal effect of interventions to allow structure deductions.
The ATE can also be estimated which is useful when interventions can be observed but not preferably applied. Common methods are the difference in difference methods [37], propensity score matching [35], and regression discontinuity designs [16].
Additionally, interventional effects can be identified by probabilistic discovery methods described in 4.2. Commonly, structure interventions can be identified with unconditional independence tests as the graph skeleton is disrupted. While soft interventions can be identified by conditional independence tests as by

adding new causes to a variable a new v-structure is created.
All in all, discovery by manipulation is a less favored discovery approach as interventions can be costly, unethical or even unfeasible for some variables. Also the strong assumption of ceteris paribus conditions are a disadvantage since some conditions are hard to reciprocate and can most often only be tackled by averaging over larger sample sets.

4.4 Structure clues from functional modeling

This approach uncovers causal relations by remodeling how the effect emerged from the cause as the cause needs to produce the effect.
So called functional causal models consist of a causal graph and a set of functions that relate the variables of the graph in accordance to the graph edges. An approach of structure discovery is to uncover the causal graph by retracing the causal functions. For example, Linear, Non-Gaussian, Acyclic causal Models (LiNGAM) [39] is an algorithm that uncovers a multivariate DAG structure by computing linear functions of the variables X and a connection strength matrix $\mathbf{B}$ plus additive noise ϵ, $X = \mathbf{B} * X + \epsilon$, by the use of an independent component analysis [17, 8]. Other methods that follow a similar approach are Additive Noise modeling, or Post Nonlinear Causal Model described in the next subsection.
These methods do not require the faithfulness assumption, but assume causal sufficiency: the absence of confounding variables. They also require the assumption that the additive noise are non-Gaussian distributions of non-zero variances, also abbreviated as *non-Gaussianity assumption*, because [42] has shown that methods that use only covariance matrices have no way of inferring the direction of the causal relation. Another fallacy of this approach is that it is prone to spurious correlations. Those may equally result in believable functions, but are not causally related. Hence the combination with another approach is strongly advised.

4.5 Structure clues from noise

To discover causal relations by the use of noise is a rather new notion. The fundamental property of causal relationships is exploited that natural noise found in the cause needs also to be found in the effect, but no noise of the effect may be found in the cause. This approach came up with Additive Noise Modeling (**ANM**), originally a functional discovery approach which also makes use of the noise property to gain certainty of the relation to be causal. Other than LiNGAM, the non-linear function $Y = f(X) + \epsilon$ of two variables X and Y is reconstructed from data. First, a regression for $X \to Y$ and $Y \to X$ is performed to approximate the relationship function f, then the residual ϵ is calculated and finally, the residuals are tested for independency [25]. This method has shown to be prone to confounding and feedback noise as both mess with the aforementioned noise property. An extension to ANMs are Post Nonlinear Causal Models (PNLs) [48] which also take nonlinear distortions f_2 from sensor or measurement errors into account by recovering $Y = f_2(f_1(X) + \epsilon)$.
Naturally, the approach of using the noise property relies on the presence of noise which is not always given. Also, it requires a large sample size of observations to reliably retrace the causal function.

4.6 Structure clues from time, space and spacetime information

Temporal and spatial information have shown to be the most important cues for human causal understanding [21]. For humans the temporal sequences of events are particularly important for the discovery of causal relationships. Events that follow a chosen event are often understood as consequences. Whereas events that precede it are understood as causes of that event [6]. Hidden in this understanding is the basic, well-known assumption that in time, a cause must always precede its effect.
Another early notion of causal understanding is the spatial proximity of the effect to its causes. In physics, this notion was called principle of locality: two objects may only be causally connected by mechanical influences as for example by touch. With the discovery of electromagnetic and gravitational waves this

notion changed. By today, we know causal relations are not only bound to space or to time but to spacetime as the travel of a causal signal is fundamentally limited by lightspeed. More current research inspects causal relations in relation to algebra in four-dimensional Minkowski spacetime M of special relativity as it can not be represented in Euclidean geometry [44]. Therein, a set of points form a region in M. An event at point x emits two light cones: a *forward lightcone* $V_+(x)$ emitted into the future, and a *backward lightcone* $V_-(x)$ emitted into the past. A point y of an event caused by x has per definition to be in $V_+(x)$, while a point of an event causing x has to be in $V_-(x)$. Any event caused by x and y has to lie within $V_+(x) \cap V_+(y)$.
The result of an intersection of a forward and a backward cone is called a *double cone*. Each double cone is causally complete, bounded, closed, and convex. The new law of locality can be derived from it: two regions in M are causally disjoint and thereby physically independent if they are spacewise separated [44]. As measurements of events in Minkowski spacetime tend to be imprecise, current literature also tackles the implications of imprecise time and location measurements and time-frame measurements [20].
The consideration of regions in Minkowski spacetime may add to the considerations in causal discovery. For example, [47] created a theoretical framework for causal image synthesis using knowledge over Minkowski spacetime.

4.7 Structure clues from forecasting and prediction

This approach assumes that if two variables are causally connected then we should be able to predict the effect given the cause. It is closely related to functional modeling, but differs in the fact that not the causing function is recovered, but instead we investigate how the prediction improves, if we add or remove knowledge over the potential cause. This approach is especially popular in timeseries. The earliest method was Granger causality for applications in the economy [13]. A stationary timeseries X is said to granger-cause another stationary timeseries Y considering X when calculating the variance of the residual of predicting Y creates a noticable change. This does not include any definition of the predicting function itself.

A derivative of Granger causality is Instantaneous causality [32] which defines X and Y as instantaneously causally related if adding the value X_i for timepoint i improves the prediction of Y_i. Countless other methods have developed in this domain like Sims causality [40] or multistep causality [23]. A common shared weakness of these methods is a sensitivity to feedback or confounders, latent variables that actually cause the observed variables, since these also allow to make predictions but are not based on a direct causal relationship.

5 Future Prospects

Future causal discovery methods will be highly dependent on the availability of data, but also on interpreting it most efficiently. Throughout the course of this paper, we highlighted several causal approaches to gain structure knowledge from all kinds of data to construct causal graphs on. We assume that efficient causal discovery algorithms will have to apply several approaches as the discovery potential of each approach is limited. With this combination of approaches, new research questions arise: 1) As each new approach is able to reduce the equivalence class of DAGs new types of equivalence classes come up which are in comparison to Markov equivalence classes of probabilistic discovery heavily underexplored. 2) New combinations of the approaches are possible which have not been investigated yet, as for example discovery methods using interventions to artificially introduce noise for discovery. For some combinations of approaches, the foundation stone is laid but still require additional work, like implementations of probabilistic algorithms that can use existing domain or expert knowledge. 3) We see a prospect in a new kind of active causal structure learning that can apply each approach most cost-efficiently for structure learning to have the highest knowledge gain [12]. 4) Common applications of causal discovery do not include technical systems, but the methods show high potential for this domain, as experiments on machines cannot be unethical and states can be easier intervened on and reciprocated [18].

6 Small Overview

We introduced the Joint Discovery Assumption which allows the joint use of multiple causal discovery methods. Also, we gave a slim overview over various notions to gain causal structure clues, i.e. the presence or absence of causal relations. For each approach, advantages, assumptions, and limits were identified. For some of the methods listed, we need additional information as temporal and or location data to make causal deductions. While in the case of expert knowledge or additional domain knowledge, the learning process requires human assistance.

Some approaches, as structure clues from spacetime, provide only structure clues over the absence of edges. This shows potential to be a cheap possibility of using additional information, while speeding up the structure search algorithms by eliminating invalid spurious relations in advance. Finally, we made a basic proposal for further research based on the presented approaches.

References

[1] Hirotogu Akaike. “Information theory and an extension of the maximum likelihood principle”. In: *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.

[2] Harish S Bhat and Nitesh Kumar. “On the derivation of the bayesian information criterion”. In: *School of Natural Sciences, University of California* 99 (2010).

[3] Giorgos Borboudakis and Ioannis Tsamardinos. “Incorporating Causal Prior Knowledge as Path-Constraints in Bayesian Networks and Maximal Ancestral Graphs”. In: *International Conference on Machine Learning* (2012).

[4] Philippe Brouillard et al. “Typing assumptions improve identification in causal discovery”. In: *International Conference on Machine Learning Workshop* (2021).

[5] John Campbell. “An interventionist approach to causation in psychology”. In: *Causal learning: Psychology, philosophy and computation* (2007), pp. 58–66.

[6] Patricia W Cheng and Laura R Novick. “A probabilistic contrast model of causal induction.” In: *Journal of personality and social psychology* 58.4 (1990), p. 545.

[7] David Maxwell Chickering. “Optimal structure identification with greedy search”. In: *Journal of machine learning research* 3 (2002), pp. 507–554.

[8] Pierre Comon. “Independent component analysis, a new concept?” In: *Signal processing* 36.3 (1994), pp. 287–314.

[9] Gregory F Cooper and Edward Herskovits. “A Bayesian method for the induction of probabilistic networks from data”. In: *Machine learning* 9.4 (1992), pp. 309–347.

[10] Frederick Eberhardt and Richard Scheines. “Interventions and causal inference”. In: *Philosophy of science* 74.5 (2007), pp. 981–995.

[11] Clark Glymour, Kun Zhang, and Peter Spirtes. “Review of Causal Discovery Methods Based on Graphical Models”. In: *Frontiers in Genetics* 10 (2019), p. 524. ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00524.

[12] Julius Gonsior et al. “Active Learning for Spreadsheet Cell Classification.” In: *EDBT/ICDT Workshops*. 2020.

[13] Clive WJ Granger. “Investigating causal relations by econometric models and cross-spectral methods”. In: *Econometrica: journal of the Econometric Society* (1969), pp. 424–438.

[14] Yangbo He, Jinzhu Jia, and Bin Yu. “Counting and exploring sizes of Markov equivalence classes of directed acyclic graphs”. In: *The Journal of Machine Learning Research* 16.1 (2015), pp. 2589–2609.

[15] Patrik Hoyer et al. “Nonlinear causal discovery with additive noise models”. In: *Advances in Neural Information Processing Systems*. Vol. 21. Curran Associates, Inc., 2009, pp. 689–696.

[16] Guido W Imbens and Thomas Lemieux. “Regression discontinuity designs: A guide to practice”. In: *Journal of econometrics* 142.2 (2008), pp. 615–635.

[17] Christian Jutten and Jeanny Herault. “Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture”. In: *Signal processing* 24.1 (1991), pp. 1–10.

[18] Andreas Kimmig et al. “Wertstromkinematik–Produktionssysteme neu gedacht”. In: *Zeitschrift für wirtschaftlichen Fabrikbetrieb* 116.12 (2021), pp. 935–939.

[19] Kevin Korb et al. “Varieties of causal intervention”. In: *Pacific Rim International Conference on Artificial Intelligence* 3157 (2004), pp. 322–331. DOI: 10.1007/978-3-540-28633-2_35.

[20] Olga Kosheleva and Vladik Kreinovich. “Observable causality implies Lorentz group: Alexandrov-Zeeman-type theorem for space-time regions”. In: 2 (30) (2014).

[21] David A Lagnado and Steven A Sloman. “Time as a guide to cause.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32.3 (2006), p. 451.

[22] Alan M. Leslie and Stephanie Keeble. “Do six-month-old infants perceive causality?” In: *Cognition* 25.3 (1987), pp. 265–288. ISSN: 0010-0277. DOI: https://doi.org/10.1016/S0010-0277(87)80006-9.

[23] Helmut Lütkepohl and Maike M Müller. “Testing for multi-step causality in time series”. In: (1994).

[24] Christopher Meek. “Causal inference and causal explanation with background knowledge”. In: *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. 1995, pp. 403–410.

[25] Joris Mooij and Dominik Janzing. “Distinguishing between cause and effect”. In: *Causality: Objectives and Assessment*. PMLR. 2010, pp. 147–156.

[26] Pramod Kumar Parida, Tshilidzi Marwala, and Snehashish Chakraverty. “A multivariate additive noise model for complete causal discovery”. In: *Neural Networks* 103 (2018), pp. 44–54.

[27] Judea Pearl. “Causal diagrams for empirical research”. In: *Biometrika* 82 (1995).

[28] Judea Pearl. *Causality*. Cambridge university press, 20090. DOI: 10.1017/CBO9780511803161.

[29] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, 1988.

[30] Judea Pearl. *The Do-Calculus Revisited*. 2012. arXiv: 1210.4852 [cs.AI].

[31] Daniel J Povinelli and Jesse M Bering. "The mentality of apes revisited". In: *Current Directions in Psychological Science* 11.4 (2002), pp. 115–119.

[32] J Michael Price. "The characterization of instantaneous causality: A correction". In: *Journal of Econometrics* 10.2 (1979), pp. 253–256.

[33] Josephine Rehak. "A Proposal on Discovering Causal Structures in Technical Systems by Means of Interventions". In: 2020, pp. 91–106.

[34] Thomas S Richardson and James M Robins. "Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality". In: *Center for the Statistics and the Social Sciences, University of Washington Series.* 128.30 (2013), p. 2013.

[35] Paul R Rosenbaum and Donald B Rubin. "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70.1 (1983), pp. 41–55.

[36] Richard Scheines et al. "The TETRAD project: Constraint based aids to causal model specification". In: *Multivariate Behavioral Research* 33.1 (1998), pp. 65–117.

[37] Hannes Schellhorn. *Effizienzeffekte der Einkommensteuer bei Steuervermeidung*. Springer-Verlag, 2005.

[38] Bernhard Schölkopf. "Causality for machine learning". In: *arXiv preprint arXiv:1911.10500* (2019).

[39] Shohei Shimizu et al. "A linear non-Gaussian acyclic model for causal discovery." In: *Journal of Machine Learning Research* 7.10 (2006).

[40] Christopher A Sims. "Money, income, and causality". In: *The American economic review* 62.4 (1972), pp. 540–552.

[41] N Sloane. *Number of acyclic digraphs (or DAGs) with n labeled nodes*. `http://oeis.org/A003024`. Accessed: 2021-11-02. 2021.

[42] Peter Spirtes et al. *Causation, prediction, and search*. MIT press, 2000. DOI: `10.1007/978-1-4612-2748-9`.

[43] Martin Stuart-Fox. "The origins of causal cognition in early hominins". In: *Biology & Philosophy* 30.2 (2015), pp. 247–266.

[44] Launey J Thomas III and Eyvind H Wichmann. "On the causal structure of Minkowski spacetime". In: *Journal of Mathematical Physics* 38.10 (1997), pp. 5044–5086.

[45] Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. "The max-min hill-climbing Bayesian network structure learning algorithm". In: *Machine learning* 65.1 (2006), pp. 31–78.

[46] Thomas Verma, Judea Pearl, et al. "Equivalence and synthesis of causal models". In: (1991).

[47] Athanasios Vlontzos et al. "Causal Future Prediction in a Minkowski Space-Time". In: *arXiv preprint arXiv:2008.09154* (2020).

[48] Kun Zhang and Aapo Hyvärinen. "Distinguishing causes from effects using nonlinear acyclic causal models". In: *Causality: Objectives and Assessment*. PMLR. 2010, pp. 157–164.