

# To do or not to do: finding causal relations in smart homes

Kanvaly Fadiga<sup>\*†</sup>, Ada Diaconescu<sup>\*</sup>, Jean-Louis Dessalles<sup>\*</sup>, Étienne Houzé<sup>\*</sup>,

<sup>\*</sup>LTCI Lab, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France

Email: {first name}.{last name}@telecom-paris.fr

<sup>†</sup>Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, France

Email: {first name}.{last name}@polytechnique.edu

**Abstract**—Research in Cognitive Science suggests that human beings understand and represent knowledge of the world through causal relationships. In addition to observations, they can rely on experimenting and counterfactual reasoning – i.e. referring to an alternative course of events – to identify causal relations and explain atypical situations. Different instances of control systems, such as smart homes, would benefit from having a similar causal model, as it would help the user understand the logic of the system and better react when needed. However, while data-driven methods achieve high levels of correlation detection, they mainly fall short of finding causal relations, notably being limited to observations only. In this paper, we propose an approach to learn causal models, combining observed data and selected interventions on the environment. We use this approach to generate Causal Bayesian Networks, which can be later used to perform diagnostic and predictive inference. We use our method on a smart home simulation, a use case where having knowledge of causal relations pave the way towards explainable systems. Our algorithm succeeds in generating a Causal Bayesian Network close to the simulation’s ground truth causal interactions, showing encouraging future prospects for application in real-life systems.

**Index Terms**—Causal Structure Discovery, Smart Home, Causal Inference

## I. INTRODUCTION

Self-Adaptive Systems (SAS) are by nature interacting with a changing environment, be it software or physical[26]. In this context of interactions, the ability to model the environment is prime, as it would help to trace back failures, identifying conflicts between goals or perform an explanatory reasoning. [9] has showed that, in typical smart home setups, explaining decisions to the user reduce the risk of wrong interventions. However, identifying causal relations in the environment of a SAS is a hard task.

Hard-coding the causal model, i.e. expressing constraints and links upon variables as rules or a static ontology is possible, but shows limited interest in the case of SAS. Indeed, since adaptability to a changing environment is a core feature of SAS, a static model of the environment is not suited to this configuration. Operating changes to the model could be considered, but is likely to require many human intervention, thus contradicting the principles of autonomic computing[6].

To illustrate this issue, consider the following situation. A user is experiencing an anomaly in the temperature control system of her smart home, as the temperature is unexpectedly

cold. A hard-coded model has been implemented, which contains causal links from heater or thermometer malfunctions to the mishandling of the temperature. However both these possibilities are discarded, as no component seems to report any problem. In this case, the cause might lie in an unexpected relation: as the user recently moved the temperature sensor closer to a light bulb, and that the days, in the winter, are shorter, the light is on, which produces heat, effectively making any measure by the thermometer wrong. This configuration being particular to this home, no hard-coded prior causal knowledge would be able to anticipate it without ad-hoc rules.

To overcome this common pitfall, many recent smart home systems integrate Machine Learning components to predict the environment’s behavior and make optimized decisions[7]. However, spurious correlations are often found in data, especially in high dimensions[2], leading to misinterpretation and erroneous causal relations. These approaches thus mostly fall short of providing the user with a comprehensible causal model of the environment.

The theory of Causality, mostly brought to the attention by J. Pearl [16, 14], brings tools to identify and eliminate spurious correlations in the construction of a causal model, mostly by formalizing the concept of *intervention*. Our method is to augment a standard Machine Learning approach with interventions on selected variables to infer causal relations. The result is a Causal Bayesian Network, i.e. a Bayesian Network whose structure is a causal graph of the environment.

Our approach is generic and can be applied to build causal models of various environments. But it can be computationally expensive to apply it to an environment with a very large network. We choose to apply it in the the smart homes case because it offers many advantages. Firstly, we don’t start from scratch as we can begin with a hard-coded causal model then incrementally improve it. Moreover, making interventions in a smart home is easier to do than in some environments (e.g. a nuclear power plant). Furthermore, the area of influence of some variables is limited to their neighborhood, which reduces the number of relationships to consider. Finally, for the scaling we can do the construction using a multi-scale approach.

The rest of this paper is organized as follows. In section II, we present the theoretical bases of causality and Bayesian graphs upon which our approach is build. Then, we review some existing approaches to related issues in section III. We

then detail our method, and propose, in section IV to illustrate it by comparing known causal graphs of an electrical circuit and a smart home with the results of our method in section V. Finally, we analyze the current limits of our approach and see how it can be integrated into broader systems in section VI.

## II. THEORETICAL BACKGROUND




Level (Symbol)	Typical Activity	Typical Question	Examples
1 Associational $P(y x)$ 	<b>Seeing</b> ML - (Un)Supervised (Bayes Net, DTree, SVM, DNN...)	What is? How would seeing X change my belief in Y?	What does seeing the light tell us about the presence of someone?
2 Interventional $P(y do(x), c)$ 	<b>Doing</b> ML - Reinforcement (Causal Bayes Net, MDPs, ...)	What if? Why if I do X?	What if I set the heater to level 5, will the temperature change?
3 Counterfactual $P(y_{-i}(x) x', y')$ 	<b>Imagining, Retrospection</b> Structural Causal Model	Why? What if I had acted differently?	Was it the heater that increase the temperature?

Fig. 1: Ladder of causation

Many examples from Machine Learning point out that algorithms usually lack the understanding of causal relations behind observations and predictions, causing misinterpretations of correlations[13]. [16] goes further by integrating this observation into a “ladder of causation”, in which three distinct levels are identified (fig. 1):

- **Observing** corresponds, according to Pearl, to the first and more reachable level of cognition: observing the world and noting correlations, dependence between some sets of variables. This stage is the ground for many modern AI approaches based on data analysis.
- **Acting** This advanced stage of cognition requires the agent to be able to act on some variables of the environment, observe the consequences and infer causal relations. The typical question answered at this point is “If I do this, what will happen next ?”
- **Counterfactual Thinking** At this point, the agent is able to conceptualize enough to be able to perform mental intervention operation on an alternative environment, and observe its evolution. According to Pearl, this level of cognitive ability is only reached by humans. The typical question would be : “What if the apple was two times as heavy ? Would it have fallen at a different speed?”

During the twentieth century, from the causal chains of Wright [27] to the integration of causal inference into Machine Learning algorithms [17, 18], research in the fields of Causality Theory aimed at formalizing the intuitive concepts of cause-consequence relations.

### A. Structural Causal Model

Causal models aim at representing the interactions between cause and consequence without ambiguity. From the definition of [17], a Structural Causal Model (SCM) contains  $C \rightarrow E$

if and only if  $C \equiv N_C$  and  $E \equiv f_E(C, N_E)$ . That is, if the cause variable  $C$  can be assigned to some specific random distribution while  $E$  can be computed from a deterministic function of  $C$  and some random noise  $N_E$ . Note that, in this definition, both the causal relation  $f_E$  and the effect noise  $N_E$  are independent from the cause  $C$ .

For more complex systems, where causal dependencies between variables may be multiple, we can use a *causal diagram* to represent an underlying structural causal model. A causal diagram (see Fig. 2) is a directed acyclic graph [16, 17, 23] that shows the causal relationships between variables. The nodes of the graph are the variables, and an edge  $(C, E)$  belongs to the graph if and only if  $C \rightarrow E$  belongs to the underlying SCM. For example, in the diagram presented in fig. 2, the arrow connecting variables *Heater* and *Temperature* ( $H \rightarrow T$ ) indicates that the temperature is causally influenced by the state of the heater. Another point of view is to consider *Temperature* as listening to the *Heater* variable to choose its value.

Compared to the more general definition of SCM, causal diagrams add the condition of being acyclic[14], encompassing the idea that causality flows in one direction only: if  $C$  has a causal influence on  $E$ , then  $E$  cannot have an influence on  $C$ . This further prevents a variable to have an influence onto itself.

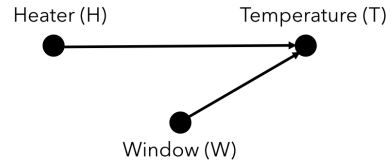


Fig. 2: A simple causal diagram.

### B. Do-operator

The idea of being able to intervene on the environment to test a causal relation between variables is prime in the literature of Causality Theory [14, 17] and can be linked to a general controlled environment experiment. As stipulated in the ladder of causation, the ability to perform this intervention operation in past observation to observe an alternative course of events is defining of human cognitive ability.

The intervention operation has been formalized by Pearl by introducing the *do-calculus*[14, 16]. Following his notations,  $do(C = x)$  means that  $C$  has been forced to take the value  $x$  by an external action. It follows that, if  $C \rightarrow E$  was part of the SCM, the causal relation  $E = f_E(C, N_E)$  remains unchanged by this operation. This operation therefore allows to identify causal relations: if  $\mathbb{P}(E) \neq \mathbb{P}(E|do(C = x))$ , there is a causal connection  $C \rightarrow E$ . In this case, we will use the notation  $do(C) \rightsquigarrow E$ .

While mere observations of the variables  $H, T$  and  $W$  from the example of fig. 2 would show correlations between  $H$  and  $T$ , interventions would give more details on the underlying SCM. On one hand,  $\mathbb{P}(Heater | do(Temperature =$

20)) =  $\mathbb{P}(Heater)$  and on the other hand,  $\mathbb{P}(Temperature | do(Heater = High)) \neq \mathbb{P}(Temperature)$ , reflecting that the heater causes the temperature change, not the other way around.

As originally stipulate, the do-operation  $do(C = x)$  considers an *external* intervention, meaning that it forces the variable  $C$  to a given value  $x$ , while making it insensitive to all other variables. On a causal diagram, this is equivalent to removing all incoming edges to node  $C$ . For instance, if we consider the simple causal diagram  $C_0 \rightarrow C_1 \rightarrow E$ , performing the intervention  $do(C_1 = x)$  will remove the edge  $C_0 \rightarrow C_1$ , thus making both  $C_1$  and  $E$  independent from  $C_0$ , thus revealing the linear structure of the graph.

### C. Bayesian Network

As Causality Theory emerged with causal diagrams, links can be made with *Bayesian Networks* which are a broadly used tool for representing and modeling correlated variables [5]. Numerous methods for training and dynamically building Bayesian Networks in many different application contexts exist in the literature[1, 5].

Formally, a Bayesian Network (see Fig. 3) is a directed acyclic graph (DAG) where the nodes correspond to random variables. Each node is associated with a set conditional probabilities  $\mathbb{P}(X_i | par(X_i))$ , where  $X_i$  is the variable associated with the specific node and  $par(X_i)$  denotes the set of parents of node  $X_i$ .

To build a Bayesian network, one therefore needs to:

- define the graph of the model, i.e. the different variables, and which ones are linked together
- find, for each of these variables, the table of probabilities conditioned on its parent variables

The graph is also called the "structure" of the model, and the probability tables its "parameters". Structure and parameters can be provided by experts, or calculated from data, although in general the structure is defined by experts and the parameters calculated from experimental data.

A Bayesian network carries no assumption that the arrow has any causal meaning. However, Bayesian networks hold the key that enables causal diagrams to interface with data. Probabilistic properties of Bayesian networks and the belief propagation algorithms that were explain later are in fact indispensable for understanding causal inference.

The main differences between Bayesian networks and causal diagrams lie in how they are constructed and the uses to which they are put. A Bayesian network is literally nothing more than a compact representation of a huge probability table. The arrows mean only that the probabilities of child nodes are related to the values of parent nodes by a certain formula (the conditional probability tables) and that this relation is sufficient. That is, knowing additional ancestors of the child will not change the formula. Likewise, a missing arrow between any two nodes means that they are independent, once we know the values of their parents.

If, however, the same diagram has been constructed as a causal diagram, then both the thinking that goes into the

construction and the interpretation of the final diagram change. In the construction phase, we need to examine each variable, say  $C$ , and ask ourselves which other variables it "listens" to before choosing its value. The chain structure  $A \rightarrow B \rightarrow C$  means that  $B$  listens to  $A$  only,  $C$  listens to  $B$  only, and  $A$  listens to no one; that is, it is determined by external forces that are not part of our model.

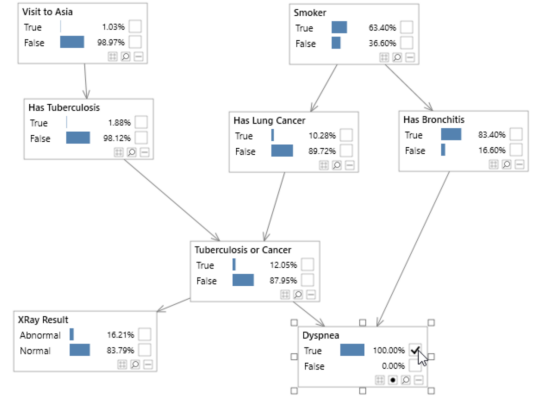


Fig. 3: A simple Bayesian network, known as the Asia network.[22]

### III. RELATED WORK

Over recent years, different approaches have tried to close the gap between "classical" observation-based Machine Learning and Causal Theory. For instance, Reinforcement Learning, as already noted by Pearl [16] can be seen as a better approach than pure correlation observations, since the agent has the opportunity to act on its environment and learn from its reactions[24]. Thus, Reinforcement Learning has proved very powerful in tasks that were previously considered as requiring intelligent thinking, such as games[21].

[12] uses a different approach to learning Bayesian Networks, trying to identify a minimal equivalence class between DAGs that fit with the observation data. The result is then presented as a Partially Directed Acyclic Graph (PDAG). While this method offers the advantage of keeping the graph simple and shows good predictive performance, it still only relies on mere observations, and as such lack causal information that may impact its interpretation. [11] try to discover the directions of the remaining edges of PDAG by means of experiment (intervention). However, the PDAG is based only on correlations, so we end up with connections based on correlation that are not causal and also missed causal relation.

Some applications consider counterfactual reasoning and integrate it into the learning process of a SCM[10, 25]. In their workflow, they consider the agent to learn a causal model of its environment then use this knowledge to perform counterfactual reasoning and improve performance. Results in providing explanations for an agent's behavior in the controlled environment of a strategy game are encouraging[10]. In a closer-to-life situation, [18] found that allowing do-operations in a learning framework could improve performance in a

classification task and achieve better-than-humans detection of medical condition.

Our approach aims at completing these encouraging steps of mixing Causality Theory and Machine Learning. Our proposed method is to learn a Causal Graph from observations and interventions on the environment, then use it as a structure to build a Bayesian Network.

#### IV. LEARNING CAUSAL BAYESIAN NETWORKS WITH INTERVENTION

The base intuition for our approach is to test whether an intervention on one variable  $C$  has an influence over other variable  $E$ , observed as a change in their distribution. If so, we know from Causality Theory that there is a causal relation  $C \rightarrow E$  in the SCM of the system, but an ambiguity remains whether this relation is direct or not. We therefore propose to incrementally block causal paths of nodes connected to the node on which we act, effectively narrowing down the possible relations.

We illustrate our approach in a setup consisting of Boolean variables. For illustration purpose, we consider the simple situation displayed in fig. 4: a room whose temperature (hot or cold) is influenced by the state of a heater (on or off). The heater can be triggered by the user’s presence in the room. Similarly, the window can be either open or close,

##### A. Testing causal influence using interventions

1) *Direct Influence:* Testing the direct influence between two variables  $C$  and  $E$  is answering the following question: “Does  $C$  have an influence on  $B$ ?”. Our approach to this question is to incrementally remove possible causal relations following different intervention. These interventions are made by directly acting upon the environment and monitoring possibly influenced variables for changes in their probability distribution. To test possible change, we use a chi-squared  $\chi^2$  test on the distributions  $\mathbb{P}(E | do(C = 0))$  and  $\mathbb{P}(E | do(C = 1))$ .

This test allows to remove non-causal connections between pairs of variables, using both intervention operation and counterfactual reasoning. The intervention operations can be performed by directly letting our algorithm acting on selected variables in the environment, thanks to the preconditions we applied on the setup. For instance, in the example of Fig. 4, the distribution of  $L$  changes depending on whether the person is detected inside ( $P=1$ ) the room or not ( $P=0$ ). Conversely, the distribution of  $P$  is not affected by the value assigned to  $L$  during the intervention operation. These two operations therefore lead to the conclusion that  $(P) \rightsquigarrow L$  is true and  $do(L) \rightsquigarrow P$  is false.

##### 2) Conditional Influence:

*Did A have influence on C given B? ( $do(A) \rightsquigarrow C | do(B)$ ):* The case of evaluating a conditional causal influence can be summarized with the question: “Did  $C$  have an influence on  $E$  given  $B$ ?” As opposed to the previous case, the causal path is indirect and thus requires additional processing. Here, we process by testing if the causal influence between  $C$  and  $E$

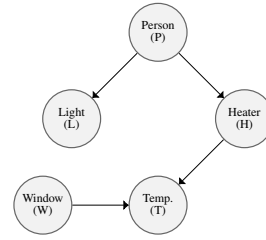


Fig. 4: A room causal diagram

still holds conditioned on the value of the third variable  $B$ . That is, checking if, for some value  $u$  taken by  $B$ , we would observe, via a chi-squared test, a difference between  $\mathbb{P}(E | do(C = 0), do(B = u))$  and  $\mathbb{P}(E | do(C = 1), do(B = u))$ .

This operation can be viewed as “locking” the value of  $B$  to a given value  $u$ , and observe if the causal relation holds. In the examples of fig. 5c and 5d, we infer the causal relations:

- $do(P) \rightsquigarrow T | do(L = 0)$  : True
- $do(P) \rightsquigarrow T | do(H = 0)$  : False

##### B. Causal Learner Algorithm

Our algorithm, presented in alg.1, iterates over the previously described elementary steps to remove non-causal pairs of variables. To this end, we start by considering a fully connected graph over all the variables in the system (see fig. 6). Then, selected causal influence tests allow to remove arrows for unrelated variables. These tests are performed by increasing order of conditioning: this allows to test the costlier high-order conditioned influences on less arrows, as many have already been discarded by the first series of tests.

As shown in fig. 6, a major limitation of this approach is that some do-operations are not feasible in realistic setups: in our example, this is the case for the temperature variable  $T$ , as one does not arbitrarily set the temperature of a room to some fixed value without modifying other variables (e.g. the heater

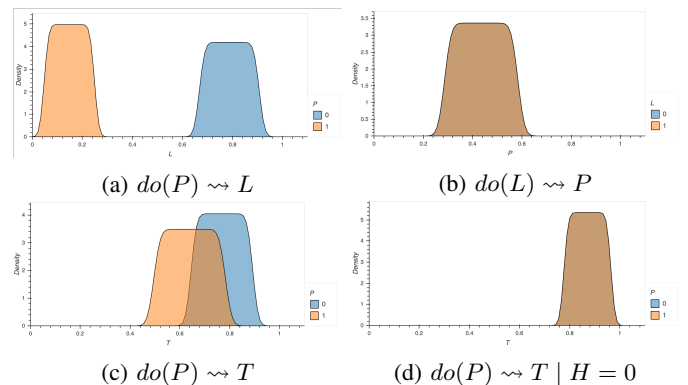


Fig. 5: Different intervention tests. (5a), the probability density of  $L$  changes depending on whether the intervention sets  $P$  to 0(blue) or 1(orange). In (5b), interventions on  $L$  do not affect the probability distribution of  $P$ . (5c): intervening on  $P$  shows a change in the distribution of  $T$ . However, conditioning this relation with  $H = 0$  removes the relation(5d)

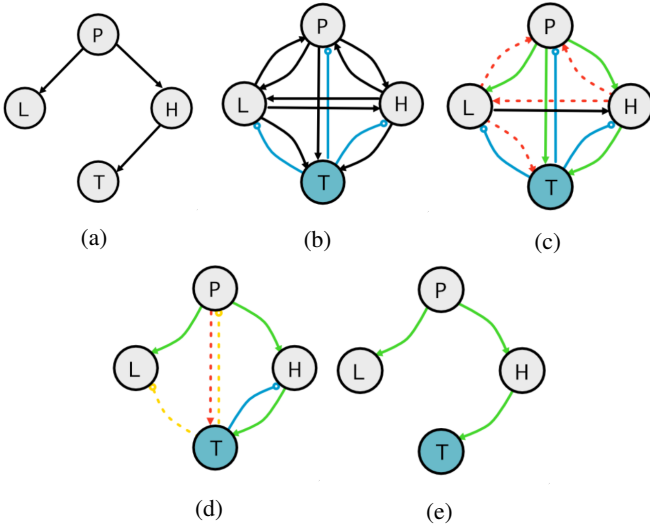


Fig. 6: Principle of our algorithm. (6a) ground truth causal model of the environment. (6b) Initialization to a fully-connected graph over the variables. Non-doable arrows and nodes are shown in blue. (6c) Causality tests with interventions remove the red arrows. (6d) Arrows are removed, either by independence test (in yellow) or causality test (in red). (6e) The final graph is obtained by removing cycles, prioritizing non-doable arrows.

state  $H$ ). We therefore call the corresponding temperature node a *non-doable node*, and consider all of the outgoing relations as “*non-doable arrows*”, or ND-arrows, in the graph. These arrows are not directly removable since the corresponding do-operations cannot be performed.

#### Algorithm 1 Extended *do* Causal Learning Algorithm

```

1: Initialization:
2:  $\mathcal{G}$  is the fully connected graph over nodes of  $\mathcal{X}$ 
3:  $k \leftarrow 0$ 
4: while There are nodes with more than  $k$  neighbors in  $\mathcal{G}$  do
5:   for each such node  $X_A$ , each of its neighbors  $X_B$  do
6:     for each subset  $\mathcal{S}$  of  $k$  neighbors of  $X_A$  do
7:       # Influence test for doable node
8:       if  $X_A$  is doable then
9:         Compute  $do(X_A) \rightsquigarrow X_B | do(\mathcal{S})$ 
10:        Remove  $A \rightarrow B$  from  $\mathcal{G}$  if need be
11:       # Independence test for non-doable node
12:       else
13:         Compute  $Corr(X_A, X_B | \mathcal{S})$ 
14:         Remove  $X_A \rightarrow X_B$  if variables are independent
15:       end if
16:     end for
17:   end for
18:    $k \leftarrow k + 1$ 
19: end while
20: Postprocess  $\mathcal{G}$  to turn it into a DAG by removing least significant arrows.

```

Processing ND-arrows therefore requires another approach. First, similarly to the PC-algorithm from [23], we use a simple chi-squared test to identify whether the two variables are

correlated, since a causal relationship implies a correlation between variables. This first step allows to remove some connections, but, for the remaining connections, it does not provide any direction for the relation. Furthermore, one needs to be cautious about the potential risk of mislabeling correlations as causal relations. As such, remaining ND-arrows should be considered only as candidate causal relations.

	Present relationships	Potential output	
1	 $A \rightarrow B$ (green) $A \rightarrow B$ (blue)	A is the cause of B Exactly one of the following holds: 1. A is the cause of B 2. B is the cause of A 3. No connection, there is a spurious correlation create by unmeasured confounder	 $A \rightarrow B$ (blue) undirected connection
3	 $A \rightarrow B$ (green) $A \rightarrow B$ (blue)	A is the cause of B	$A \rightarrow B$ (green)
4	 $A \rightarrow B$ (red dashed) $A \rightarrow B$ (blue)	A <b>is not</b> the cause of B. Exactly one of the following holds: 1. B might be the cause of A 2. No connection, there is a spurious correlation create by unmeasured confounder	 $A \rightarrow B$ (blue) Flagged connection
5	 $A \rightarrow B$ (blue) $C \rightarrow A$ (green) $C \rightarrow B$ (green)	The confounder is known, so we can delete the spurious correlation create by C.	 $C \rightarrow A$ (green) $C \rightarrow B$ (green)

Fig. 7: The different possible configurations for processing the remaining ND-arrows. ND-arrows are shown in green, regular ones in blue.

To handle the rest of the process, we rely on the fact that the causal diagram is, by definition, a DAG. This condition leads to the removal of some arrows among the remaining candidates. Depending on the configuration, different possibilities are considered, as fig. 7 shows:

- **case 1:** As no ambiguity exists, the arrow is kept in the graph.
- **case 2:** In this case, no information can be gathered through correlation study alone. If no direction creates a cycle in the graph, the algorithm will keep the undirected relation, and tag it as potentially spurious. Further data may lead to eliminating both of the arrows.
- **case 3:** Here, one direction of the relation has been tested through a do-operation, while the other has not. The algorithm will therefore keep the direction that has been tested with an intervention.
- **case 4:** While this ND-arrow can be a spurious correlation, the algorithm will keep it if it does not create a cycle in the resulting graph. It will however be flagged as such. Otherwise, the arrow is removed from the graph. More generally, if keeping several ND-arrows would lead to a cycle, the algorithm will remove the least significant one with respect to Chi-square score.
- **case 5:** In this case we see an ND-arrow that can be preserved if there is a confounder that creates a correlation between A and B. Here we see that C is a confounder. So we drop the ND-arrow.

After processing all ambiguous cases, the algorithm outputs a DAG representing a causal model compatible with observations from the system. This causal diagram can be

be used as a basis for further analysis. A first possibility is to use it to infer potential causes to unusual situations, and as such be included into a broader-scoped explanation process[3]. A second prospect, detailed here, is to use this diagram as the structural basis of a Bayesian Network for finer causal inference.

### C. Causal Model to Bayesian Network

In the literature, training a Bayesian is usually divided into two main parts[5] : learning the structure of the graph and estimating its parameters. Since we use the previously learned causal diagram as a base structure, we will only focus in this part on learning the different parameters of the network, i.e. the probability tables for each node conditioned on its parents. We will call the resulting graph a *Bayesian Causal Network* to emphasize its particular structure: while usual Bayesian Networks do not entail causality between their nodes, our approach leads to a graph whose connections entail a cause-effect relation.

To estimate these parameters, a conventional approach is to use a maximum likelihood estimator[**TODO**], which can be resumed as estimating variables values given their parents' values only from past observational data. For example, if we consider the graph from fig. 8, we would compute  $p_{00}$  with:

$$p_{00} = \frac{N_{T=0,(W,H)=(0,0)}}{N_{T=0,(W,H)=(0,0)} + N_{T=1,(W,H)=(0,0)}} \quad (1)$$

where  $N_{T=i,(W,H)=(j,k)}$  is the number of past occurrences of  $(T, W, H) = (i, j, k)$ .

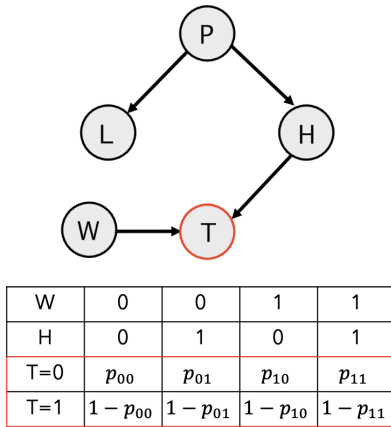


Fig. 8: Example of a small Bayesian network. The probability table for node T is displayed.

However this conventional approach might be face some issues for some estimations, notably if the number of occurrences is small. For instance, in our small example of fig. 8, it might not be possible to estimate the parameters  $p_{01}$ . The introduction of the do-operation removes this limitation, as it becomes possible, for doable nodes, to generate all kinds of situations required to observe the outcome and estimate missing parameters of the Bayesian Network.

### D. Causal Inference on Causal Bayesian Network

Upon completion of the training, the resulting Bayesian Network can be seen as a “conditional probability machine”[5]. It can be used for different tasks requiring inferring new knowledge on the system. For instance, [19] shows how a Bayesian Network can be used to compute the probabilities of different diseases compatible with the observed symptoms. The inference can then also be used to infer probabilities of yet unseen symptoms and which further examinations would be most useful. This example shows the different possibilities offered by a Bayesian Network: diagnostic and predictive inference.

- **Predictive:** This kind of inference is interested in “guessing” the most probable future state of the system, given a configuration, i.e. answering the question: “*What happens if X is equal to x?*” As fig. 9a shows, if evidence is put on node P, the inference will propagate following the direction of causal arrows, to the children of the affected node P.
- **Diagnostic:** On the other hand, diagnostic inference is interested in looking into the probable causes of observed consequences: “*what would be the cause of X = x?*” The inference therefore goes backwards, as displayed in fig. 9b: from the observation on L, we infer the probable state of P, which will entail consequences over H and T.

In either case, inference works as follows: we denote by  $Bel(X = x)$  the belief that a node takes a given value (see fig. 3, where beliefs are displayed for each node). Following an observation of the system, the beliefs of one or several nodes are set to a set value. For instance, in fig. 8, knowing that the person is present will set the value of P to 1 with probability 1. This change to beliefs will then be propagated through the graph, following Bayes’ rule.

While we let the details of the propagation algorithm out of the scope of this paper (readers interested in a complete description of the process may refer to [15, 5]), we could visualize the propagation mechanism as follows.

The propagation algorithm is iterative. At every step, each node X passes the following messages: to its children Y,  $\pi_X(Y)$  containing transition probabilities; to its parents Z,  $\lambda_X(Z)$  containing likelihood information. Conversely, it receives messages  $\pi_Z(X)$  from its parents, and  $\lambda_Y(X)$  from its children (see fig.9). Each node then updates its beliefs according to the messages it receives:

$$Bel(X) = \alpha \lambda(X) \pi(X) \quad (2)$$

where  $\alpha$  is a normalizing factor,  $\lambda(X) = \prod_Y \lambda_Y(X)$  and  $\pi(X) = \prod_Y \pi_Y(X)$  are the products of all messages received from children and parents, respectively. As shown by [15], for DAGs, this propagation method converges to the beliefs values satisfying the observations and the network’s parameters in a finite number of steps.

Predictive and diagnostic inference then allows to answer various queries about the environment without having to further intervene on the system. Applications of such knowledge

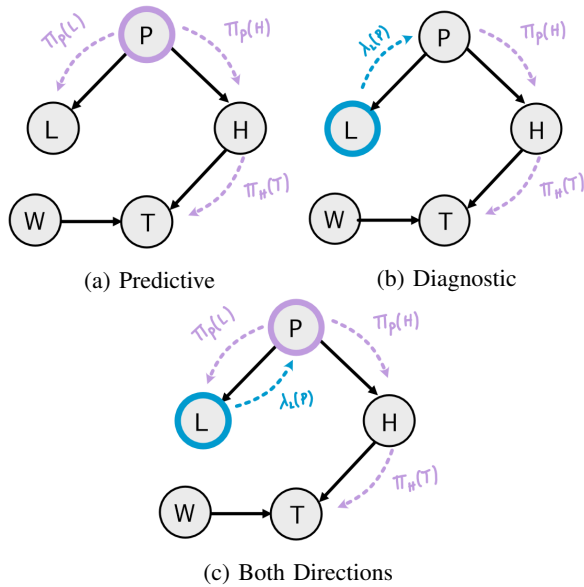


Fig. 9: Belief propagation in a Bayesian Network can be either forwards (9a) for predictive applications, backwards (9b) for diagnostic purpose, or both-oriented (9c).

is further discussed in sec. VI. One might however note that, as opposed to a traditional Bayesian Network, our proposed CBN uses only causal relations. As such, one might argue that the entailed reasoning appears more “natural”, a case confirmed by the observation that causal relations are algorithmically simpler [17, 4].

	( $x=0, x=1$ )
P	(0.5, 0.5)
L	(0.45, 0.55)
H	(0.7, 0.3)
T	(0.67, 0.33)
W	(0.7, 0.3)

TABLE I: Prior probability of each node of 9b

	( $x=0, x=1$ )
P	(0.89, 0.11)
L	(1, 0)
H	(0.86, 0.14)
T	(0.73, 0.27)
W	(0.7, 0.3)

TABLE II: Posterior probability after  $L=0$ .

## V. EXPERIMENTS AND RESULTS

### A. General Workflow

As previously stated in sec. I, we apply our methods to a smart home environment. This choice is motivated by various reasons. First, smart homes provide good examples of closed environments monitored by SAS. As such, they also provide simulators, which can be used to implement an intervention operator without being limited by common physical constraints (time, safety issues, incompatibilities). In addition, they can present unusual or surprising situations where the use of a causal diagnostic can help intervene on the system to improve performance[9, 3]. The following section describes in detail our workflow, from data generation to training the CBN and using it for inference tasks.

Boolean variable	Simulation measure
person	(User.x, User.y) $\in$ room
thermometer	room.temperature $\leq$ threshold
outdoor	outdoor.temperature $\leq$ threshold
light	light.powerStatus = 1
presence	sensor.presenceSensed = 1
power	house.powerConsumption $\leq$ threshold
thermometer	thermometer.temperature $\leq$ threshold
window	window.open = 1

TABLE III: Correspondence between simulation measures and the Boolean variables.

1) *Smart Home simulator*: Our experiments are built upon the iCasa platform[8]. Based on the OSGi framework, it offers a service-oriented platform for simulations of smart home physical systems. Its autonomic manager keeps track of currently used devices, which allows for runtime deployment and modification of configuration. This enables the simulation of scenarios where variable interactions are more intricate. In our example, we simulated the behavior of different rooms, each one characterized by physical variables such as temperature, illumination. Each of these rooms is equipped with different devices able to monitor or modify the room’s physical variables: heater, thermometer, presence sensor, light, etc. Table III shows the different monitored variables of the example. The entire configuration is shown in fig. 10 using the iCasa Web UI.

Using a simulation, as opposed to using a real setup, brings two main advantages for our experiments. First, it allows to have a perfect knowledge of the groundtruth causal interactions, as they are directly encoded into the simulator. Secondly, it provides an easy control over different parameters and thus allow to perform, if desired, some interventions that would not be feasible in real-life. This will allow to test the effect of having access to more or less possible interventions for our algorithm.



Fig. 10: The iCasa GUI showing the basic setup for our experiments: four rooms equipped with a presence sensor triggering heating and lighting, and a thermometer.

2) *Observation Data Generation*: Once the initial setup is complete, we let the simulation run while the different house’s devices are left in “autonomous mode”, i.e. they are able to adapt to changes of condition to maintain some key

environment variables within a target range, for instance the temperature and CO2 concentration of each room. At runtime, we randomly act on some of these variables or components to observe how the system reacts to change. In total, our continuous observation generated around 500 data points. Since values from variables are originally numerical, we convert them to Boolean value by using simple threshold comparisons. Thus, we obtain a set of Boolean observations which are used to observe correlations between variables.

3) *Intervention Data Generation:* To perform interventions, we use the possibility offered by the simulator setup to disable some devices. Disabled devices will no longer react to their input sensors, thus achieving the Markov blanket independence implied by intervention[14]. Then the value of the device is set to a fixed value. For instance, the intervention  $do(Heater = 1)$  will cause the heater to turn on while being insensitive to any environment factor such as the detection of the user's presence.

To generate intervention data, we then proceed as follows: we sample the house in a state  $s$  where each variable is assigned a value, and from this state, make an intervention  $do(X = x)$  on a selected variable. We did 20 interventions per node at each stage. After a set time period  $\Delta_t$ , we measure the resulting state of the house  $s'$ , eventually considering only variables of interest (variables correlated to  $X$ ). The period  $\Delta_t$  is set to a given value manually chosen from prior experiments with the simulator, to allow the system reach an equilibrium state after the intervention. We will discuss further time considerations in sec. VI.

### B. Results

The causal model of the simulation we used for our experiments is displayed in fig. 11. We first consider three variables, namely the presence sensor, the house's power consumption and the room's temperature as ND-nodes. While the simulation setup would allow us to intervene on them, we introduced this limitation to observe the impact of ND-nodes on causal discovery.

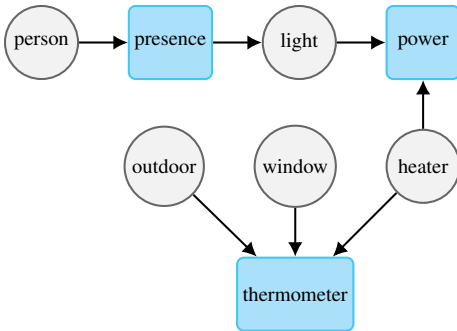


Fig. 11: Groundtruth causal model for the living room. Non-doable variables are shown in blue.

1) *Causal Bayesian Network learning:* The construction of the causal graph is shown in fig. 13. First, observations of correlated variables and results of interventions yields a “raw”

output depicted in fig 12a. Note how the presence of ND-nodes introduces ND-arrows emerging from them. The next step of our algorithm processes this raw output to remove the least significant arrows to make it a DAG that is compatible with the observations. The output of this step, shown in fig. 12b, contains two arrows flagged as ND-arrows. When comparing this final output to ground truth, in fig. 13c, we notice that one of these ND-arrows was erroneous, displaying a performance limit in the case of ND-nodes. On the other hand, one causal relation, between light status and power power consumption, was missed by our approach. This mistake can be explained in this situation, by the relatively small impact of light, in comparison to the heater, on power consumption.

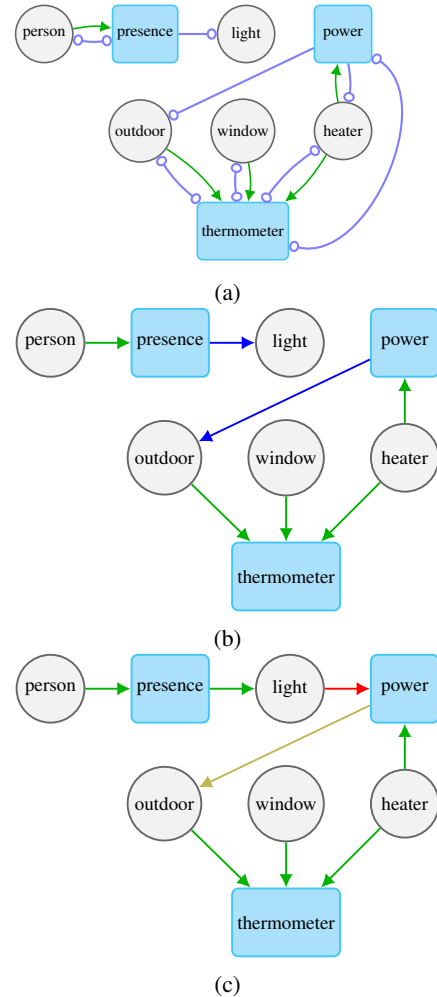


Fig. 12: Results of our approach applied to the smart home simulation. The raw output of conditional testings, shown in (12a), is then processed to remove less significant arrows to obtain a DAG (12b). (13c): comparison between this output and the ground truth diagram from fig. 11: the red arrow is a missed relation while the yellow one is a connection wrongly added to the model.

Building on the structure of the causal graph presented in fig. 13, we complete the learning process by using maximum



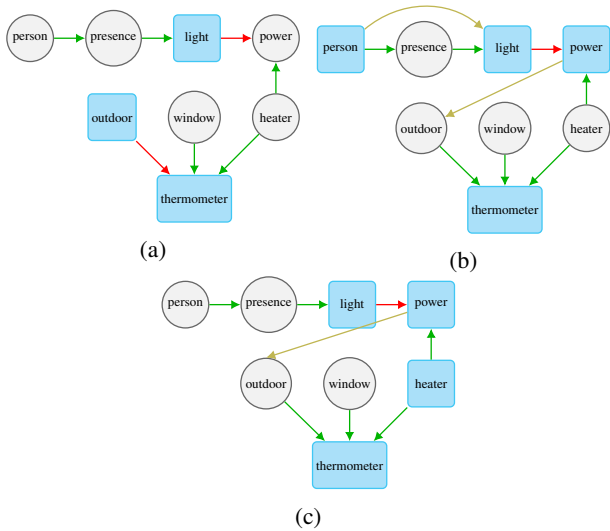


Fig. 13: Results of different setup of ND-node compare with the ground truth diagram from fig. 11

likelihood estimates to finally provide a Causal Bayesian Network. Table IV shows the probability table for the thermometer, conditioned by its three parents nodes: heater, outdoor and window.

heater	0	0	0	0	1	1	1	1
outdoor	0	0	1	1	0	0	1	1
window	0	1	0	1	0	1	0	1
thermometer=0	0.5	X	0	0	1.0	1.0	X	X
thermometer=1	0.95	X	1.0	1.0	0	0	X	X

TABLE IV: Conditional probability table of thermometer.

## 2) Changing Learning Setup:

3) *Unusual Causal Relations*: A motivation for learning CBN with minimal prior knowledge was the ability to adapt to unusual situations, detect them and use knowledge to provide explanation to the user. Such a situation may occur in our experimental setup: the thermometer has been implemented to be sensitive to the heat produced by nearby devices (notably the light or the heater). In our setup, four rooms are simulated with the same devices, however the precise location of devices within each room is random. This leads to situations where, in one of the rooms, the light is sufficiently close to the light so that a new causal relation ( $light \rightarrow thermometer$ ) appears in the causal model of the room. Face with this event, our algorithm was able to see the new connection in the corresponding room.

In this particular case, the diagnostic inference offered by our approach allows to find the cause of a peculiar behavior of the system: “*The thermometer reports a hot temperature while the heater is off*”. In this case, diagnostic inference on the Bayesian network would initiate  $T = 1$  and  $H = 0$ , and would infer  $P(L = 1 | (T, H) = (1, 0)) = \mathbf{TODO}$ .

## VI. DISCUSSION AND FUTURE WORK

While our experimental setup of the smart home was set to be close to a real-life use case, some limitations still remain in our approach, some of which will be discussed here.

The first major assumption is that the causal model of the system can always be represented by an acyclic graph. This limitation is common in the literature on Causality Theory ([17, 14, 23]). However, especially in the context of SAS, it is likely that retro-actions occur between devices and the variables they monitor: for instance, consider how a “smart” heater would turn on depending on the room’s temperature. One potential workaround would be to take time into account, which removes any ambiguity regarding the direction of a causal relation.

However adding time to the equation is no easy task: in sec. V, we argued that for the purpose of our demonstration, we used a fixed time period  $\Delta_t$  after which we consider the consequences of interventions. This fixed value entails several issues: as discussed in [TODO], different causal mechanisms can have different time characteristics; how can one knows how long is long enough when waiting for the consequences of an intervention? Existing methods propose to estimate the time interval following interventions [todo]: in the near future, we consider integrating a similar approach to the learning process of our Causal Bayesian Networks, as to reduce the number of parameters.

Furthermore, our approach requires a certain number of interventions on the system, and has shown to perform better when only a limited number of variables are non-doable. These issues are minor in a simulated example, but can be limiting when operating on a real-life environment. A workaround is to consider to have access to a model on the environment, such as a “digital twin”[20], which our algorithm can use.

Having causal diagram of a system, as opposed to a simple Bayesian Network, offers possibilities to be integrated into explanations frameworks. For instance, we may use Causal Bayesian Network in conjunction with the general Explanatory Engine proposed by [3]. This use case would benefit from both inference directions: diagnostic can be used as a powerful tool to propose hypotheses for abductive inference (i.e. finding the cause of an observed phenomenon), while predictive inference might be used to explore the potential consequences of a proposed solution, in the context of an explanation.

Future work may focus on optimising learning for large network, for example, use hierarchical learning, i.e. learning the causal network of several areas and then learn the causal relationships between them. Or find a way to reduce the number of do-operations.

## VII. CONCLUSION

We work towards the implementation in real-life SAS of the methods of Causality Theory. We have seen how intervention operations could be performed on a digital twin of an environment to train a causal diagram, which can later be used as a basis for our Causal Bayesian Network. This workflow has shown encouraging results in the example of

a smart home and, since it required no ad hoc knowledge about the particularities of the smart home context, can be generalized to other comparable setups.

Knowing the Causal Bayesian Graph offers advantages for applications such as explanations, given the more “natural” source of relations it entails, compared to a more traditional Bayesian Network. As such, we consider using this tool as a mean to perform abductive and predictive inference in a broader explanatory framework.

#### ACKNOWLEDGMENT

The authors would like to thank...

#### REFERENCES

- [1] Concha Bielza and Pedro Larranaga. “Discrete Bayesian network classifiers: A survey”. In: *ACM Computing Surveys (CSUR)* 47.1 (2014). Publisher: ACM New York, NY, USA, pp. 1–43.
- [2] Jianqing Fan, Fang Han, and Han Liu. “Challenges of big data analysis”. In: *National science review* 1.2 (2014). Publisher: Oxford University Press, pp. 293–314.
- [3] Etienne Houzé et al. “A Decentralized Approach to Explanatory Artificial Intelligence for Autonomic Systems”. In: *ACSOS 2020 Conference Proceedings, Companion*. Aug. 2020.
- [4] Dominik Janzing and Bernhard Schölkopf. “Causal inference using the algorithmic Markov condition”. In: *IEEE Transactions on Information Theory* 56.10 (2010). Publisher: IEEE, pp. 5168–5194.
- [5] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [6] Jeff Kramer and Jeff Magee. “A rigorous architectural approach to adaptive software engineering”. In: *Journal of Computer Science and Technology* 24.2 (2009), pp. 183–188.
- [7] J Naveen Ananda Kumar and Srija Chimmani. “Proposal of smart home resource management for waste reduction and sustainability using AI and ML”. In: *2019 International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2019, pp. 992–998.
- [8] Philippe Lalanda and Catherine Hamon. “A service-oriented edge platform for cyber-physical systems”. In: *CCF Transactions on Pervasive Computing and Interaction* 2.3 (2020). Publisher: Springer, pp. 206–217.
- [9] Nianyu Li et al. “Explanations for human-on-the-loop: A probabilistic model checking approach”. In: *Proceedings of the IEEE/ACM 15th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*. 2020, pp. 181–187.
- [10] Prashan Madumal et al. “Explainable reinforcement learning through a causal lens”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. Issue: 03. 2020, pp. 2493–2500.
- [11] Stijn Meganck, Philippe Leray, and Bernard Manderick. “Learning Causal Bayesian Networks from Observations and Experiments: A Decision Theoretic Approach”. In: *Modeling Decisions for Artificial Intelligence*. Ed. by Vicenç Torra et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 58–69. ISBN: 978-3-540-32781-3.
- [12] Bojan Mihaljević, Concha Bielza, and Pedro Larrañaga. “Learning Bayesian network classifiers with completed partially directed acyclic graphs”. In: *International Conference on Probabilistic Graphical Models*. PMLR, 2018, pp. 272–283.
- [13] Ziad Obermeyer et al. “Dissecting racial bias in an algorithm used to manage the health of populations”. In: *Science* 366.6464 (2019). Publisher: American Association for the Advancement of Science, pp. 447–453.
- [14] Judea Pearl. *Causality: Models, Reasoning and Inference*. 2nd. USA: Cambridge University Press, 2009. ISBN: 0-521-89560-X.
- [15] Judea Pearl. “Fusion, propagation, and structuring in belief networks”. In: *Artificial intelligence* 29.3 (1986). Publisher: Elsevier, pp. 241–288.
- [16] Judea Pearl and Dana Mackenzie. *The Book of Why. The New Science of Cause and Effect*. New York: Basic Books, 2018. ISBN: 978-0-465-09760-9.
- [17] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.
- [18] Jonathan G Richens, Ciarán M Lee, and Saurabh Johri. “Improving the accuracy of medical diagnosis with causal machine learning”. In: *Nature communications* 11.1 (2020). Publisher: Nature Publishing Group, pp. 1–9.
- [19] Pedro Pereira Rodrigues et al. “Causality assessment of adverse drug reaction reports using an expert-defined Bayesian network”. In: *Artificial intelligence in medicine* 91 (2018). Publisher: Elsevier, pp. 12–22.
- [20] Roland Rosen et al. “About the importance of autonomy and digital twins for the future of manufacturing”. In: *IFAC-PapersOnLine* 48.3 (2015). Publisher: Elsevier, pp. 567–572.
- [21] Julian Schrittwieser et al. “Mastering atari, go, chess and shogi by planning with a learned model”. In: *Nature* 588.7839 (2020). Publisher: Nature Publishing Group, pp. 604–609.
- [22] Bayes Server. *Asia Network*. URL: <https://www.bayesserver.com/examples/networks/asia>.
- [23] Peter Spirtes et al. *Causation, prediction, and search*. MIT press, 2000.
- [24] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [25] Sergei Volodin, Nevan Wichers, and Jeremy Nixon. “Resolving spurious correlations in causal models of environments via interventions”. In: *arXiv preprint arXiv:2002.05217* (2020).

- [26] Danny Weyns. “Software Engineering of Self-adaptive Systems”. In: *Handbook of Software Engineering*. Ed. by Sungdeok Cha, Richard N. Taylor, and Kyochul Kang. Cham: Springer International Publishing, 2019, pp. 399–443. ISBN: 978-3-030-00262-6. DOI: 10.1007/978-3-030-00262-6\_11. URL: [https://doi.org/10.1007/978-3-030-00262-6\\_11](https://doi.org/10.1007/978-3-030-00262-6_11).
- [27] Sewall Wright. “Correlation and causation”. In: *Journal of agricultural Research* 20 (1921), pp. 557–580.