Supplementary Material

1 EXPRESSING A STRUCTURAL CAUSAL MODEL AS AN EXTENDED STRUCTURAL CAUSAL MODEL

Algorithm 1 has as input a *Structural Causal Model* (SCM) and as output an *Extended Structural Causal Model* (ESCM). An ESCM is a richer model than a SCM in that the likelihoods over interventions are specified in ESCM and not in SCM, therefore, an algorithm that outputs an ESCM and takes as input an SCM should also, either implicitly or explicitly, take as input the information $\mathcal{P}(\mathbf{D}|\mathbf{U})$. The input $\mathcal{P}(\mathbf{D}|\mathbf{U})$ is present in the algorithm to make it clear that some information is added in the transformation, however is marked as optional in ESCM as every query answerable by a SCM (by asserting that an intervention took place) is answerable by an ESCM without the usage of the information in $\mathcal{P}(\mathbf{D}|\mathbf{U})$.

In Figure S1 a graphical representation of a 5 variable SCM is presented alongside with two graphical representations for a corresponding ESCM. The graphical representation of Figure S1.a is based on a *Causal Bayesian Network* (CBN), the graphical representation of Figure S1.b is based on *Bayesian Network* (BN) and the graphical representation of Figure S1.c is based on *Markov Network* (MN). The distinction between CBN and BN is necessary as the structure in the graphical representation in Figure S1.a contains information that is used by an algorithm external to the SCM to adjust it in the case of interventions by replacing variable definitions given by functions, that is $\mathbf{v_i} \triangleq f_i(Pa(\mathbf{v_i}, \mathcal{F}))$. Two different structures with the same input-output relations, given by different composition of the functions in $\mathcal{F}$, can yield two different answers to interventional queries posed to a SCM. This is not the case in the representation used in Figure S1.b where all input-output relationships including interventions are defined explicitly, making the underlying graphical model a BN. In Figure S1.c the MN structure is obtained by moralization of the BN of Figure S1.b.

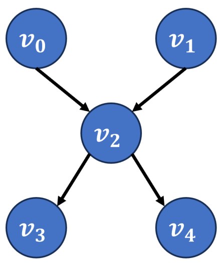

Figure S1.a. Directional SCM representation.

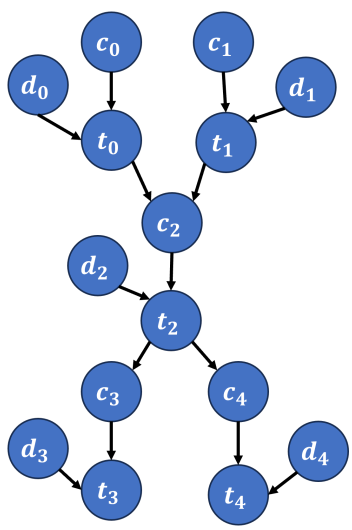

Figure S1.b. Directional ESCM representation.

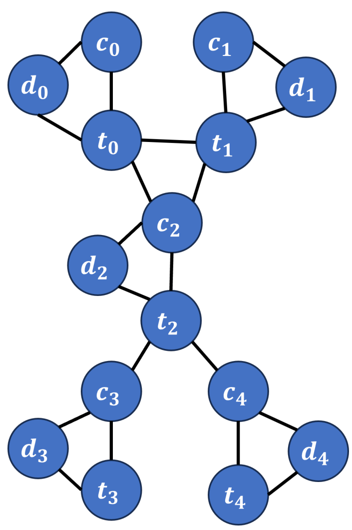

Figure S1.c. Non directional ESCM representation.

Figure S1. Example of SCM and ESCM representations.

Algorithm 1 Construction of an ESCM from a SCM.

Input SCM defined by the tuple $\big(\mathbf{U_{SCM}}, \mathbf{V}, \mathcal{F}_{SCM}, \mathcal{P}(\mathbf{U_{SCM}})\big)$;
Input (Optional): $\mathcal{P}(\mathbf{D}|\mathbf{U})$
Output: ESCM defined by the tuple $\big(\mathbf{U}, \mathbf{C}, \mathbf{T}, \mathbf{D}, \mathcal{F}, \mathcal{G}, \mathcal{P}(\mathbf{U}, \mathbf{D})\big)$.

Create empty sets of variables $\mathbf{U}, \mathbf{C}, \mathbf{D}, \mathbf{T}$
Create empty sets of functions $\mathcal{F}, \mathcal{G}$
$\mathbf{U} = \mathbf{U_{SCM}}$
for all $\mathbf{v_i} \in \mathbf{V}$ **do**
    Create $\mathbf{c_i}$
    Add to the variable $\mathbf{c_i}$ the set of states that the variable $\mathbf{v_i}$ has
    Add $\mathbf{c_i}$ to $\mathbf{C}$
    Create $\mathbf{t_i}$
    Add to the variable $\mathbf{t_i}$ the set of states that the variable $\mathbf{v_i}$ has
    Add $\mathbf{t_i}$ to $\mathbf{T}$
    Create $\mathbf{d_i}$
    Add to the variable $\mathbf{d_i}$ the state corresponding to "no intervention" ▷ $\cancel{d}_i^{\emptyset}$, see the notation section
    Add to the variable $\mathbf{d_i}$ a state for each state of the variable $\mathbf{v_i}$ has. Each of this states has the meaning of intervention that sets the variable $\mathbf{t_i}$ to the respective value in the set of values a $\mathbf{v_i}$ can take. ▷ $\forall value \in Val(\mathbf{v_i})$: add $\cancel{d}_i^{value}$ to the states of $\mathbf{d_i}$, see the notation section
    Add $\mathbf{d_i}$ to $\mathbf{D}$
end for
for all $f_{SCM_i} \in \mathcal{F}_{SCM}$ **do**
    Create f_i
    Define f_i as a copy of f_{SCM_i} where in the inputs each occurrence of $\mathbf{v_j} \in Pa(\mathbf{v_i}, \mathcal{F})$ is replaced with a corresponding $\mathbf{t_j}$ and the output ($\mathbf{v_i}$) is replaced with a corresponding $\mathbf{c_i}$
    Add f_i to the set $\mathcal{F}$
    Create g_i that takes as inputs $\mathbf{c_i}$, $\mathbf{d_i}$ and outputs $\mathbf{t_i}$
    $g_i(\mathbf{c_i}, \mathbf{d_i} = \cancel{d}_i^{\emptyset}) = \mathbf{c_i}$
    $g_i(\mathbf{c_i}, \mathbf{d_i} = \cancel{d}_i^{value}) = value \; \forall value \neq \emptyset \in Val(\mathbf{d_i})$
    Add g_i to the set $\mathcal{G}$
end for
if $\mathcal{P}(\mathbf{D}|\mathbf{U})$ is not defined in the input **then**
    $\mathcal{P}(\mathbf{D}|\mathbf{U}) \leftarrow$ Uniform Distribution
end if
$\mathcal{P}(\mathbf{U}, \mathbf{D}) \leftarrow \mathcal{P}(\mathbf{D}|\mathbf{U}) \times \mathcal{P}(\mathbf{U_{SCM}})$ ▷ In SCM the likelihoods of interventions are not defined so a usage of an ESCM for conditional queries is the same as for a SCM, as the conditioning of the queries on the interventions erases the information added in this step
return the tuple $\big(\mathbf{U}, \mathbf{C}, \mathbf{T}, \mathbf{D}, \mathcal{F}, \mathcal{G}, \mathcal{P}(\mathbf{U}, \mathbf{D})\big)$.

Consider that every variable in a SCM, called $Example_{5Var}$ with the structure of Figure S1.a is binary, that is: $\forall\ 0 \leq i \leq 4\ \{v_i, \overline{v_i}\} = Val(\mathbf{v_i})$. In that SCM, $\mathbf{v_0} \xrightarrow{\text{Causes}} \mathbf{v_2}$, $\mathbf{v_1} \xrightarrow{\text{Causes}} \mathbf{v_2}$, $\mathbf{v_2} \xrightarrow{\text{Causes}} \mathbf{v_3}$ and $\mathbf{v_2} \xrightarrow{\text{Causes}} \mathbf{v_4}$, so: $\mathbf{v_0} \triangleq f_0(\mathbf{U})$, $\mathbf{v_1} \triangleq f_1(\mathbf{U})$, $\mathbf{v_2} \triangleq f_2(\mathbf{v_0}, \mathbf{v_1}, \mathbf{U})$, $\mathbf{v_3} \triangleq f_3(\mathbf{v_2}, \mathbf{U})$ and $\mathbf{v_4} \triangleq f_4(\mathbf{v_2}, \mathbf{U})$. Let's consider that a distinct subset of variables in $\mathbf{U}$ is an argument in each of the functions in $\mathcal{F}$. Let's also consider that for each plausible function ($f_{i_j}(Pa(\mathbf{v_i}, \mathcal{F}))$) that provides a value to a variable $\mathbf{v_i}$ there exists a variable $\mathbf{U_i}$ such that:1) $f_0(\mathbf{U_0})$ is defined according to equation S1; 2) $f_1(\mathbf{U_1})$ is defined according to equation S2; 3) $f_3(\mathbf{v_2}, \mathbf{U_3})$ is defined according to equation S3; 4) $f_4(\mathbf{v_2}, \mathbf{U_4})$ is defined according to equation S4. The definition of the function $f_2(\mathbf{v_0}, \mathbf{v_1}, \mathbf{U_3})$ has 16 different[1] values for

[1] The number of mappings form a set $\mathbf{A}$ to a set $\mathbf{B}$ ($B \leftarrow A$) is $|B|^{|A|}$.

$\mathbf{U_2}$:$\{u_{2_{FFFF}}, ..., u_{2_{TTTT}}\}$ where the subscript has a representation of the output of the respective function given a combination of values of the remaining inputs (similarly to the rest of the functions). The exogenous variables ($\mathbf{U}$) completely specify the underlying function that yielded a variable its value. The uncertainty over the states of $\mathbf{U}$ ($\mathcal{P}(\mathbf{U})$) is reflected in uncertainty about the definition of each of the functions $f_i \in \mathcal{F}$. The values $P(\mathbf{U_0} = u_{0_T}) = 0.2$ and $P(\mathbf{U_0} = u_{0_F}) = 0.8$ yield $P(v_0) = 0.2$ and $P(\overline{v_0}) = 0.8$. Uncertainty in $\mathbf{U_3}$ signals not only uncertainty about the value of $\mathbf{v_3}$ but also uncertainty about the role $\mathbf{v_2}$ takes in the assignment of a value to $\mathbf{v_3}$. An intervention in an SCM changes the function that defines a variable and so impacts the relationship among variables $\mathbf{U}$ and $\mathbf{V}$ as all functions remain functions of $\mathbf{U}$ but a subset of variables $\mathbf{V}$ stand for different expression in $\mathbf{U}$. This changes all functions that have those expressions as inputs.

Applying algorithm 1 to $Example_{5Var}$, for each $\mathbf{v_i}$ with $\{v_i, \overline{v_i}\} = Val(\mathbf{v_i})$ we get a $\mathbf{c_i}$ with $\{c_i, \overline{c_i}\} = Val(\mathbf{c_i})$, a $\mathbf{t_i}$ with $\{t_i, \overline{t_i}\} = Val(\mathbf{t_i})$ and a $\mathbf{d_i}$ with $\{\cancel{d}_i^{\emptyset}, \cancel{d}_i^{F}, \cancel{d}_i^{T}\} = Val(\mathbf{d_i})$. For each function f_i in a SCM there exists a new function f_i in an ESCM. The variables $\mathbf{c_i}$, $\mathbf{t_i}$ in ESCM have the same states than the variable $\mathbf{v_i}$ in SCM so in place of v_i and $\overline{v_i}$, in functions f_0 through f_4, t_i and $\overline{t_i}$ should be written, e.g:f_4 in ESCM would be written in the form expressed in equation S5. The functions f_0 and f_1 would remain the same as they don't reference in their inputs any variable in the set $\mathbf{V}$. In ESCM the output of each function f_i is attributed to the corresponding $\mathbf{c_i}$, therefore we have that: $\mathbf{c_i} = f_i(Pa(\mathbf{c_i}, \mathcal{F}))$. The set of functions $g_i \in \mathcal{G}$ for ESCM is defined in equation S6. Each function g_i was defined without using a variable $\mathbf{U}$ in its input so as to capture the concept of intervention with minimal changes. Nevertheless, in ESCM the object $\mathcal{P}(\mathbf{U}, \mathbf{D})$ was used so as to avoid making claims about the relationship between any variable in the set $\mathbf{D}$ and $\mathbf{U}$. In order to set $\mathcal{P}(\mathbf{U}, \mathbf{D})$ with $\mathcal{P}(\mathbf{U})$ it is necessary to add information pertaining to $\mathcal{P}(\mathbf{D}|\mathbf{U})$ as $\mathcal{P}(\mathbf{U}, \mathbf{D}) = \mathcal{P}(\mathbf{D}|\mathbf{U}) \times \mathcal{P}(\mathbf{U})$. In the absence of such information, with the only goal of performing some computation defined for a SCM with an ESCM, any value for $\mathcal{P}(\mathbf{D}|\mathbf{U})$ can be assumed, as SCM does not use that information.

$$\begin{cases} f_0(\mathbf{U_0} = u_{0_T}) & = True \\ f_0(\mathbf{U_0} = u_{0_F}) & = False \end{cases} \tag{S1}$$

$$\begin{cases} f_1(\mathbf{U_1} = u_{1_T}) & = True \\ f_1(\mathbf{U_1} = u_{1_F}) & = False \end{cases} \tag{S2}$$

$$\begin{cases} f_3(\overline{v_2}, \mathbf{U_3} = u_{3_{FF}}) & = False \\ f_3(v_2, \mathbf{U_3} = u_{3_{FF}}) & = False \\ f_3(\overline{v_2}, \mathbf{U_3} = u_{3_{FT}}) & = False \\ f_3(v_2, \mathbf{U_3} = u_{3_{FT}}) & = True \\ f_3(\overline{v_2}, \mathbf{U_3} = u_{3_{TF}}) & = True \\ f_3(v_2, \mathbf{U_3} = u_{3_{TF}}) & = False \\ f_3(\overline{v_2}, \mathbf{U_3} = u_{3_{TT}}) & = True \\ f_3(v_2, \mathbf{U_3} = u_{3_{TT}}) & = True \end{cases} \tag{S3}$$

$$\begin{cases} f_4(\overline{v_2}, \mathbf{U_4} = u_{4_{FF}}) & = False \\ f_4(v_2, \mathbf{U_4} = u_{4_{FF}}) & = False \\ f_4(\overline{v_2}, \mathbf{U_4} = u_{4_{FT}}) & = False \\ f_4(v_2, \mathbf{U_4} = u_{4_{FT}}) & = True \\ f_4(\overline{v_2}, \mathbf{U_4} = u_{4_{TF}}) & = True \\ f_4(v_2, \mathbf{U_4} = u_{4_{TF}}) & = False \\ f_4(\overline{v_2}, \mathbf{U_4} = u_{4_{TT}}) & = True \\ f_4(v_2, \mathbf{U_4} = u_{4_{TT}}) & = True \end{cases} \tag{S4}$$

$$\begin{cases} f_4(\overline{t_2}, \mathbf{U_4} = u_{4_{FF}}) & = False \\ f_4(t_2, \mathbf{U_4} = u_{4_{FF}}) & = False \\ f_4(\overline{t_2}, \mathbf{U_4} = u_{4_{FT}}) & = False \\ f_4(t_2, \mathbf{U_4} = u_{4_{FT}}) & = True \\ f_4(\overline{t_2}, \mathbf{U_4} = u_{4_{TF}}) & = True \\ f_4(t_2, \mathbf{U_4} = u_{4_{TF}}) & = False \\ f_4(\overline{t_2}, \mathbf{U_4} = u_{4_{TT}}) & = True \\ f_4(t_2, \mathbf{U_4} = u_{4_{TT}}) & = True \end{cases} \tag{S5}$$

$$g_i(\mathbf{c_i}, \mathbf{d_i}) = \begin{cases} \mathbf{c_i} \; if \; \mathbf{d_i} = \cancelto{\emptyset}{d_i} \\ True \; if \; \mathbf{d_i} = \cancelto{T}{d_i} \\ False \; if \; \mathbf{d_i} = \cancelto{F}{d_i} \end{cases} \tag{S6}$$

2 DIRECTIONALITY, STRUCTURAL CAUSAL MODELS AND EXTENDED STRUCTURAL CAUSAL MODELS

A SCM necessarily has a *Direct Acyclic Graph* (DAG) structure. The DAG is used as an input to a procedure that adapts the SCM structure to handle interventions. The structure in equations is the main reason for the usage of the operator defined as ($\triangleq$) in SCM as a change in a definition of a variable affects the expression of its effects[2]. An ESCM does not require a DAG structure because all information we need to use is in its input-output mapping. This does not prevent us to use a DAG to convey a function decomposition that exists due to a set of cause-effect relationships. This is the case in the graphical model representation in Figure S1.b and in a *Computation Graph* (CG) that encodes the composition of all functions in the sets $\mathcal{F}$ and $\mathcal{G}$ in an ESCM. A CG that contains composition of functions is a DAG. Contrary to SCM, in ESCM we are not bound to a specific function composition, e.g: the MN expressed in S1.c can be compiled into a *Tractable Probabilistic Model* (TPM) with variable eliminations by any order. Not requiring the usage of a specific DAG structure in the computations done by ESCM enables us to replace any DAG for another that has the same input-output mapping and in that sense the ESCM is not directional.

[2] E.g: $\mathbf{v_a} \xrightarrow{\text{Causes}} \mathbf{v_b} \xrightarrow{\text{Causes}} \mathbf{v_c}$, in the absence of intervention let $\mathbf{v_a} \triangleq f_a(\mathbf{U_a})$,$\mathbf{v_b} \triangleq f_b(f_a(\mathbf{U_a}), \mathbf{U_b})$ and $\mathbf{v_c} \triangleq f_c(f_b(f_a(\mathbf{U_a}), \mathbf{U_b}), \mathbf{U_c})$. After an intervention setting $\mathbf{v_b}$ to $true$, that is replacing its definition with $\mathbf{v_b} \triangleq true$ we have, that $\mathbf{v_c} \triangleq f_c(true, \mathbf{U_c})$ and $\mathbf{v_a} \triangleq f_a(\mathbf{U_a})$. The intervention acted differently weather a variable was a cause or an effect of the intervened variable. The operator defined as ($\triangleq$) is a way to differentiate causes from effects, allowing us to communicate that they should be treated differently.

SCM were, in part, motivated by the use of causality in human language. The functional description of ESCM provides a principled way of justifying cause-effect relationships as a way of dealing with known unknowns under limited resources. Consider that:

1. Full information about a variable i ($\mathbf{t_i}$) is obtained by combining the information we get from other variables declared to be their causes ($\mathbf{c_i}$) and additional information that we know to exist ($\mathbf{d_i}$);
2. The information that is present in $\mathbf{t_i}$, but not in $\mathbf{c_i}$, is considered to be due to known unknowns ($\mathbf{d_i}$);
3. There are limited resources preventing us from using every source of information in the computation of $\mathbf{c_i}$;
4. Adding variables to the arguments of a function f_i allows its output $\mathbf{c_i}$ to better approach $\mathbf{t_i}$;
5. There are sources of information that are considered more useful than others for the estimation of the value of a variable.

This can motivate the choice of "cause" variables and the usage of cause-effect relationships in the context of ESCM as long as they are found useful. In this context, ESCM can be used without necessarily imposing a constraint on a data generating process requiring asymmetric[3] relations among variables.

3 FURTHER EMPIRICAL TESTS

3.1 Data

The empirical tests in this section are based on the Earthquake(Korb and Nicholson, 2010) dataset to which interventions over the variables a) "earthquake" ($\mathbf{d_0}$), b) "alarm" ($\mathbf{d_2}$), and c) "MaryCalls" ($\mathbf{d_3}$) were added according to the cause-effect relationships expressed in Figure S2.a. Each of the variables in the set $\mathbf{D}$ has three possible values, one corresponding to the absence of intervention over the corresponding endogenous variable and two corresponding to setting the value of the intervened variable to either of the values it can take. Existence and type of interventions were determined independently for each variable in $\mathbf{D_{0,2,3}}$. To the absence of intervention for each variable was assigned probability 50%. The likelihood over the rest of the states[4] of the variables $\mathbf{D_{0,2,3}}$ was determined so that interventions replaced the probability distribution over the states of the corresponding variable that depended on the modeled causes by a value[5] sampled from a uniform distribution. 50000 samples were created and a random split was used for separating the training data (80%) from the test data (20%).

3.2 Models and Training

Similarly to the experiments on the main paper, two types of models were used: 1) $Type_{Ord}$ expressed in equation S7 and Figure S2.b and 2) $Type_{NN,Tree}$ expressed in equation S8 and Figure S2.c . Similarly to the experiments in the main article, three instances of the second type of model ($Type_{NN,Tree}$) with different numbers of sum nodes on each L_* layer ($2n$ with n=1,2 and 4) were used. The $Type_{NN,Tree}$ models used for the Earthquake dataset have 28, 80 and 256 parameters, hence they are called $Tree_{28}$, $Tree_{80}$ and $Tree_{256}$. The $Type_{Ord}$ model used for the Earthquake dataset has 84 parameters of which 40 are non zero, hence the model was called Ord_{40}. Contrary to the case of the models in the main article, this was not the model with least parameters.

[3] A superset of directional.

[4] Corresponding to one of the possible interventions over the variable.

[5] Hard interventions that set a variable to one of its possible values were used. No soft interventions that set a variable to a distribution of its values were used.

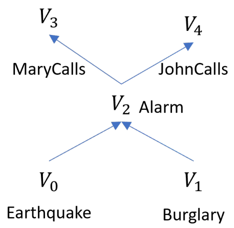

Figure S2.a. Cause-Effect Relationships in Earthquake dataset(Korb and Nicholson, 2010). Arrows point from causes to effects.

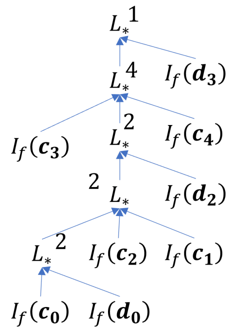

Figure S2.b. Structure of a model given by equation S7. The numbers next to L_* represent the number of $\oplus$ nodes the layer has.

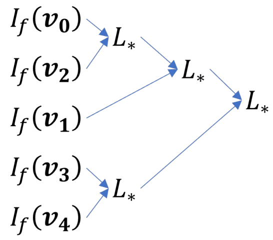

Figure S2.c. Structure of a model given by equation S8.

Figure S2. Decompositions of functions according to the ground truth (Figure S2.a), and equations S7 (Figure S2.b) and S8 (Figure S2.c).

$$L_*\big(L_*(L_*(L_*(L_*(I_f(\mathbf{C_0}), I_f(\mathbf{D_0})), I_f(\mathbf{C_1}), I_f(\mathbf{C_2})), I_f(\mathbf{D_2})), I_f(\mathbf{C_3}), I_f(\mathbf{C_4})), I_f(\mathbf{D_3})\big) \tag{S7}$$

$$L_*\big(L_*(L_*(I_f(\mathbf{V_0}), I_f(\mathbf{V_2})), I_f(\mathbf{V_1})), L_*(I_f(\mathbf{V_3}), I_f(\mathbf{V_4}))\big) \tag{S8}$$

The training was similar to that of the experiments connected using the datasets Asia and Synthetic and detailed in the main article.

3.3 Results and Discussion

The results are shown in Figures S3 to S6 where the height of each bar stands for the mean of a value over the repetitions of the experiments and the error bar has the height of two standard deviations over the repetitions of the experiments. In Figure S3 the absolute value of the average of the logarithm of the likelihood of observing the data in the test dataset given that the intervention took place is plotted. This value is minimized during training (for the training dataset that is drawn from the same statistical distribution as the test dataset) and a lower value corresponds to better modeling the data. The Ord_{40} model has better fitness and the bigger the $Type_{NN,Tree}$ model the better the fitness. The difference between the models is smaller than in the experiments made in the main paper. This is explained by the simpler relation among the variables. In Figure S4 differences in responses to queries in for the different models can be observed. In Figure S5 it can be seen that the value of $\mathbf{c_4}$ is conditionally independent on $\mathbf{d_0}$ given $\mathbf{d_2}$. There are minimal differences that are explained by the usage of random sampling in data generation and finite dataset size. In Figure S6 it is shown that it cannot be expected that $Type_{NN,Tree}$ generalizes according to what would be expected due to our knowledge of the data generating process if we don't provide it either in the data, in a set of constraints that the model should satisfy or in the objective function used in training.

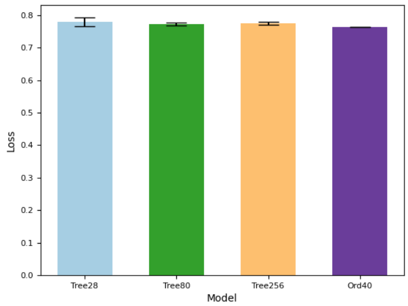

Figure S3. Absolute value of the average of logarithm of conditional likelihood of observing the test data given that the respective intervention took place. Lower is better.

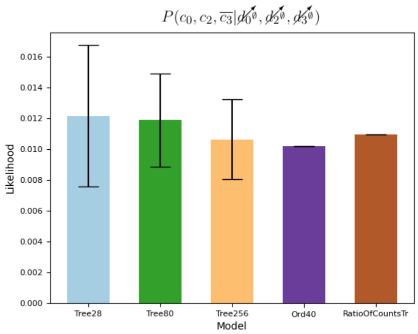

Figure S4. Response to query $P(c_0, c_2, \overline{c_3} | \cancel{d_0}^{\emptyset}, \cancel{d_2}^{\emptyset}, \cancel{d_3}^{\emptyset})$.

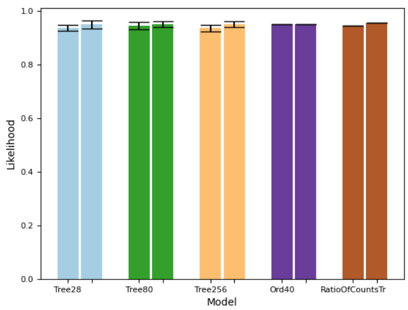

Figure S5. Response to queries $P(c_4 | \cancel{d_0}^{F}, \cancel{d_2}^{T}, \cancel{d_3}^{\emptyset})$ (left) and $P(c_4 | \cancel{d_0}^{T}, \cancel{d_2}^{T}, \cancel{d_3}^{\emptyset})$ (right).

4 SYNTHETIC DATASET

All $\mathbf{V}$ variables are binary, therefore each $P(\overline{v_i} | val_j)$ can be computed as $1 - P(v_i | val_j)$ for any $\mathbf{v_i} \in \mathbf{V}$ and any $val_j \in Val(Pa(\mathbf{v_i}, \mathcal{F}))$. It should be noted that the CBN specifies the weights when no interventions happen, that is, they are independent of the interventions that influence the distributions of the variables in the dataset.

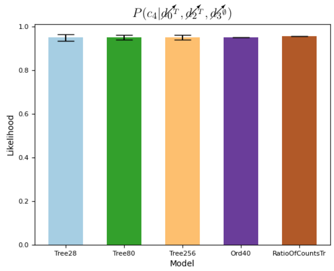

Figure S6.a. Response to query $P(c_4|\cancel{d_0}^{T}, \cancel{d_2}^{T}, \cancel{d_3}^{\emptyset})$.

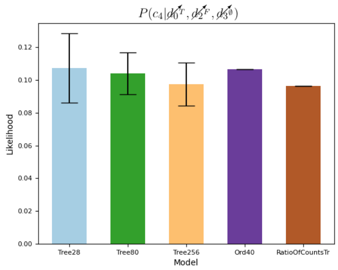

Figure S6.b. Response to query $P(c_4|\cancel{d_0}^{T}, \cancel{d_2}^{F}, \cancel{d_3}^{\emptyset})$.

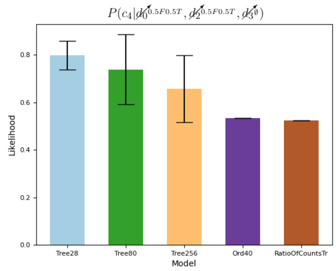

Figure S6.c. Response to query $P(c_4|\cancel{d_0}^{0.5F0.5T}, \cancel{d_2}^{0.5F0.5T}, \cancel{d_3}^{\emptyset})$.

Figure S6. Extrapolation from queries not observed during training.

The weights used for generating the data are listed below:

0.$P(v_0) = 0.9$;
1.$P(v_1) = 0.0$;
2.$P(v_2) = 0.23076923076923078$;
3.$P(\mathbf{V_3}|\mathbf{V_0}, \mathbf{V_1})$
  a.$P(v_3|v_0, v_1) = 0.4375$
  b.$P(v_3|\overline{v_0}, v_1) = 1.0$
  c.$P(v_3|v_0, \overline{v_1}) = 0.8125$
  d.$P(v_3|\overline{v_0}, \overline{v_1}) = 0.38461538461538464$
4.$P(\mathbf{V_4}|\mathbf{V_1})$
  a.$P(v_4|v_1) = 0.8888888888888888$
  b.$P(v_4|\overline{v_1}) = 0.6086956521739131$
5.$P(\mathbf{V_5}|\mathbf{V_2}, \mathbf{V_3}, \mathbf{V_4})$
  a.$P(v_5|v_2, v_3, v_4) = 0.5625$
  b.$P(v_5|\overline{v_2}, v_3, v_4) = 0.42105263157894735$
  c.$P(v_5|v_2, \overline{v_3}, v_4) = 0.0$
  d.$P(v_5|\overline{v_2}, \overline{v_3}, v_4) = 1.0$
  e.$P(v_5|v_2, v_3, \overline{v_4}) = 0.34782608695652173$
  f.$P(v_5|\overline{v_2}, v_3, \overline{v_4}) = 0.6875$
  g.$P(v_5|v_2, \overline{v_3}, \overline{v_4}) = 0.5555555555555556$
  h.$P(v_5|\overline{v_2}, \overline{v_3}, \overline{v_4}) = 0.4782608695652174$
6.$P(\mathbf{V_6}|\mathbf{V_3}, \mathbf{V_5})$
  a.$P(v_6|v_3, v_5) = 0.38461538461538464$
  b.$P(v_6|\overline{v_3}, v_5) = 1.0$
  c.$P(v_6|v_3, \overline{v_5}) = 1.0$
  d.$P(v_6|\overline{v_3}, \overline{v_5}) = 0.4$

REFERENCES

Korb, K. B. and Nicholson, A. E. (2010). *Bayesian artificial intelligence* (CRC press)