# The Hidden Elegance of Causal Interaction Models

Silja Renooij[1(✉)] and Linda C. van der Gaag[1,2]

[1] Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
s.renooij@uu.nl

[2] Dalle Molle Institute for Artificial Intelligence, Lugano, Switzerland
linda.vandergaag@idsia.ch

**Abstract.** Causal interaction models such as the noisy-OR model, are used in Bayesian networks to simplify probability acquisition for variables with large numbers of modelled causes. These models essentially prescribe how to complete an exponentially large probability table from a linear number of parameters. Yet, typically the full probability tables are required for inference with Bayesian networks in which such interaction models are used, although inference algorithms tailored to specific types of network exist that can directly exploit the decomposition properties of the interaction models. In this paper we revisit these decomposition properties in view of general inference algorithms and demonstrate that they allow an alternative representation of causal interaction models that is quite concise, even with large numbers of causes involved. In addition to forestalling the need of tailored algorithms, our alternative representation brings engineering benefits beyond those widely recognised.

**Keywords:** Bayesian networks · Causal interaction models · Maintenance robustness

## 1 Introduction

The use of *causal interaction models* has become popular as a technique for simplifying probability acquisition upon building Bayesian networks for real-world applications. These interaction models essentially impose specific patterns of interaction among the causal influences on an effect variable, by means of a *parameterised conditional probability table* for the latter variable. The number of parameters involved in this table typically is linear in the number of causes involved, where the full table itself is exponentially large in this number. Various different causal interaction models have been designed for use in Bayesian networks, the best known among which are the (leaky) noisy-OR model and its generalisations (see for example [4,11,17]).

While a causal interaction model describes a conditional probability table for the effect variable in a causal mechanism by a linear number of parameters, most software packages for inference with the embedding Bayesian network

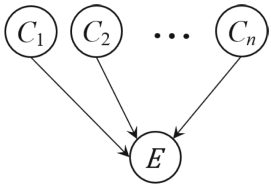
© Springer Nature Switzerland AG 2019
N. Ben Amor et al. (Eds.): SUM 2019, LNAI 11940, pp. 38–51, 2019.
https://doi.org/10.1007/978-3-030-35514-2_4

require the fully specified table. This full probability table is then generated from the parameters and the definition of the interaction model used, prior to the inference. Using fully expanded probability tables is associated with two serious disadvantages, however. Firstly, the size of the full table is exponential in the number of cause variables involved in a causal mechanism, which induces both the specification size of the network and the runtime complexity of inference to increase substantially. Secondly, using full tables has the engineering disadvantage that the modelling decision to impose a specific pattern of causal interaction is no longer explicit in the representation, as a consequence of which the intricate dependencies between the cells of the table are effectively hidden.

For richly-connected Bayesian networks with large numbers of cause variables per effect variable, as found for example from probabilistic relational models [7], inference scales poorly and quickly becomes infeasible. Over the last decades therefore, researchers have addressed ways to ameliorate the representational and inferential complexity of using fully expanded probability tables with causal interaction models. One such approach has focused on the design of tailored inference algorithms for noisy-OR Bayesian networks, which trade off general applicability and runtime efficiency; these algorithms in essence exploit the structured specification of the noisy-OR model for all variables upon inference (see for example [5,6,8,12,15]). While experimental results underline their scalability for noisy-OR networks, these tailored algorithms are not easily integrated with current algorithms for probabilistic inference in general. Another approach to tackling the representational and inferential complexity of using fully expanded probability tables for causal interaction models, has focused on the design of more concise representations of causal mechanisms; these alternative representations in essence are distilled automatically from the interaction models at hand and allow use of general inference algorithms (see for example [9,10,16,18,19]).

In this paper we reconsider and integrate some of the early work in which causal mechanisms with interaction models are represented by alternative graphical structures and probability tables. We demonstrate that interaction models with specific decomposition properties can be represented efficiently by an alternative structure with associated small tables that have an intuitively appealing semantics. This alternative structure can be readily embedded in a general Bayesian network and thereby allows for inference without the necessity of preprocessing tables or using tailored algorithms. We further argue that this alternative representation induces elegant properties from an engineering perspective which allow more ready maintenance and safer fine-tuning of parameters than the use of fully expanded probability tables in causal mechanisms.

The paper is organised as follows. In Sect. 2, we briefly review causal interaction models, and the (leaky) noisy-OR model more specifically. In Sect. 3, we reconsider the partition of causal interaction models into a deterministic function and associated independent noise variables, and demonstrate when and how the underlying deterministic function can be decomposed. Based on these insights, we derive our alternative cascading representation and study its properties in Sect. 4. We conclude the paper in Sect. 5.




| $\mathbf{c}$ | $\Pr(e \mid \mathbf{c})$ |
|---|---|
| $\bar{c}_1, \bar{c}_2, \bar{c}_3$ | $0$ |
| $c_1, \bar{c}_2, \bar{c}_3$ | $p_1$ |
| $\bar{c}_1, c_2, \bar{c}_3$ | $p_2$ |
| $\bar{c}_1, \bar{c}_2, c_3$ | $p_3$ |
| $c_1, c_2, \bar{c}_3$ | $1-(1-p_1)\cdot(1-p_2)$ |
| $c_1, \bar{c}_2, c_3$ | $1-(1-p_1)\cdot(1-p_3)$ |
| $\bar{c}_1, c_2, c_3$ | $1-(1-p_2)\cdot(1-p_3)$ |
| $c_1, c_2, c_3$ | $1-(1-p_1)\cdot(1-p_2)\cdot(1-p_3)$ |

**Fig. 1.** A causal mechanism $\mathcal{M}(n)$ with $n$ cause variables $C_i$ and the effect variable $E$ (*left*); a conditional probability table imposed by the noisy-OR model, for $n = 3$ (*right*).

## 2 Preliminaries

We briefly review causal interaction models for Bayesian networks and thereby introduce our notational conventions. In this paper, we focus on binary random variables, which are denoted by (possibly indexed) capital letters $X$. The values of such a variable $X$ are denoted by small letters; more specifically, we write $\overline{x}$ and $x$ to denote absence and presence, respectively, of the concept modelled by $X$. (Sub)sets of variables are denoted by bold-face capital letters $\mathbf{X}$ and their joint value combinations by bold-face small letters $\mathbf{x}$; $\Omega(\mathbf{X})$ is used to denote the domain of all value combinations of $\mathbf{X}$. We further consider joint probability distributions Pr over sets of variables, represented by a Bayesian network.

Within Bayesian networks, we consider *causal*[1] *mechanisms* $\mathcal{M}(n)$ composed of a single effect variable $E$ and one or more cause variables $C_i$, $i = 1, \ldots, n$, with arcs pointing to $E$; Fig. 1 (*left*) illustrates the basic idea of such a mechanism. For the effect variable $E$ of a causal mechanism, a conditional probability table is specified, with distributions $\Pr(E \mid \mathbf{C})$ over $E$ for each joint value combination $\mathbf{c}$ for its set $\mathbf{C}$ of cause variables; this table thus specifies a number of distributions that is exponential in the number of cause variables involved.

A *causal interaction model* for a causal mechanism $\mathcal{M}(n)$ takes the form of a parameterised probability table for the effect variable involved. The noisy-OR model [17], which is the best known among these interaction models, defines the conditional probability table for the effect variable $E$ of $\mathcal{M}(n)$ through

- the conditional probability $\Pr(e \mid \bar{c}_1, \ldots, \bar{c}_n) = 0$;
- the *parameters* $p_i = \Pr(e \mid \bar{c}_1, \ldots, \bar{c}_{i-1}, c_i, \bar{c}_{i+1}, \ldots, \bar{c}_n)$, for all $i = 1, \ldots, n$;
- the *definitional rule* $\Pr(e \mid \mathbf{c}) = 1 - \prod_{i \in I_{\mathbf{c}}}(1 - p_i)$ for the probabilities given the remaining value combinations $\mathbf{c}$ involving the presence of two or more causes, where $I_{\mathbf{c}}$ is the set of indices of the present causes $c_i$ in $\mathbf{c}$.

Figure 1 (*right*) illustrates the parameterised table of the noisy-OR model for a mechanism with three cause variables. For a causal mechanism $\mathcal{M}(n)$, the model

[1] Although we do not make any claim with respect to causal interpretation, we adopt the terminology commonly used.

defines a full probability table over $n+1$ variables, specifying a total of $2{\cdot}2^n$ probabilities; half of these are derived from $\Pr(e \mid \mathbf{c}) + \Pr(\overline{e} \mid \mathbf{c}) = 1$ and, hence, are redundant. Of the $2^n$ non-redundant probabilities, the noisy-OR model allows the values of only the $n$ parameter probabilities $p_i$ to be chosen freely. The model further forces the distribution $\Pr(E \mid \overline{c}_1, \ldots, \overline{c}_n)$ to be degenerate.

Since its introduction, the noisy-OR model has given rise to several variants and generalisations (see [4] for an overview). Of these, we briefly review here the *leaky noisy*-OR *model*. This model differs from the noisy-OR model in that it includes an additional *leak parameter* $p_L = \Pr(e \mid \overline{c}_1, \ldots, \overline{c}_n)$ that captures the probability of the effect $e$ occurring in the absence of all modelled causes. Different interpretations of the noisy-OR parameters in view of this leak probability have given rise to different definitional rules for the remaining probabilities [4,11]. Without loss of generality, we adopt in this paper the interpretation proposed by Díez [4], and use the rule $\Pr(e \mid \mathbf{c}) = 1 - (1-p_L) \cdot \prod_{i \in I_{\mathbf{c}}} (1-p_i)$ for the probabilities given arbitrary joint value combinations $\mathbf{c}$ with multiple present causes, where $I_{\mathbf{c}}$ again is the set of indices of the causes present in $\mathbf{c}$.

## 3 Decomposition of Causal Interaction Models

Causal interaction models are often viewed as combining a deterministic function $f$ with independent noise variables $Z_i$ per cause variable (see for example [10,14,17]); Fig. 2 (*left*) illustrates this view for the (leaky) noisy-OR model. The noise variables $Z_i$ are associated with the probabilities $\Pr(z_i \mid c_i) = p_i$, $\Pr(z_i \mid \overline{c}_i) = 0$, where the $p_i$ are the model's parameters; in the leaky variant of the noisy-OR model, the prior probability $\Pr(z_L) = p_L$ for the designated noise variable $Z_L$ is the leak parameter. The deterministic function $f$ equals the logical OR and is encoded in the probability table $\Pr(E \mid \mathbf{Z})$ for the effect variable $E$ through degenerate distributions. The variable $E$ thereby is a deterministic variable and, by convention, is indicated by a double border in our figure. Slightly abusing notation, we will further write $E = f(\mathbf{Z})$.

The representation in Fig. 2 (*left*) was introduced originally to indicate how a causal interaction model could ease the task of knowledge acquisition for causal mechanisms involving large numbers of variables [9]: by making independence of the causal influences explicit, the partition into a deterministic part and a probabilistic noise part underlines the requirement of actually just a limited number of parameters. While indeed easing the task of knowledge acquisition for practical applications, the partition of a causal interaction model does not reduce the actual size of its representation for use with general inference algorithms. In fact, embedding the partition of a causal mechanism $\mathcal{M}(n)$ in a Bayesian network will increase the total number of variables involved by $n$ and still require the specification of exponentially many probabilities for the effect variable $E$.

Specific types of causal interaction model however, actually do allow a reduced representation [10]. More formally, it are specific decomposability properties of the deterministic function $f$ that provide for a reduction of the size of the conditional probability table(s) for the effect variable(s) in a causal mechanism.

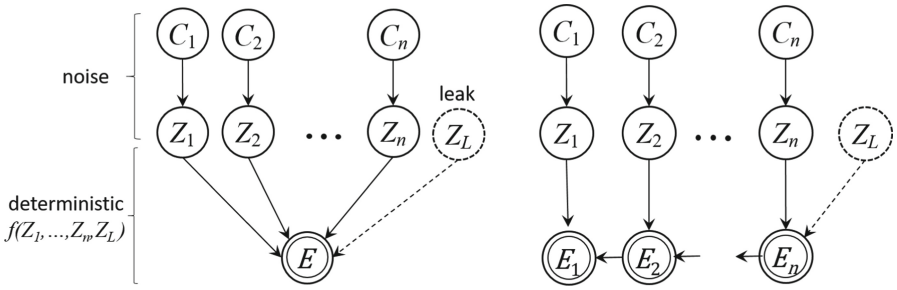



**Fig. 2.** Partition of a causal interaction model into a probabilistic noise part and a deterministic functional part (*left*); a chain decomposition for a commutative and associative deterministic function (*right*).

Such decomposability properties of functions are widely used in mathematics and computing science to simplify functions by their hidden structure: a function $f(\cdot)$ on a set of entities is called *self-decomposable* if, for any two disjoint subsets $\mathbf{X}, \mathbf{Y}$, the property $f(\mathbf{X} \cup \mathbf{Y}) = f(\mathbf{X}) \diamond f(\mathbf{Y})$ holds, for some commutative and associative merge operator $\diamond$ (cf. [13]). Commutative and associative logical operators, such as AND and OR, are self-decomposable Boolean functions. Now, if the deterministic function $f$ modelled for the effect variable $E$ in the partition in Fig. 2 (*left*) is self-decomposable, it can be split into a sequence of function applications, each to a subset of $E$'s cause variables. Each such application can then be described by an auxiliary effect variable $E_i$ with fewer parents than $E$. The set of auxiliary variables resulting from such a functional decomposition can be organised in various different graphical structures. In this paper the chained organisation from Fig. 2 (*right*) will be used and referred to as a *chain decomposition*. We would like to note that the idea of introducing additional variables to reduce the number of parents for a variable is a general modelling technique for Bayesian networks, known as *parent divorcing* [16].

We consider again the partition of a causal interaction model into a probabilistic part with noise variables $Z_i$, $i = 1, \ldots, n$, and a deterministic part $E = f(Z_1, \ldots, Z_n)$ for some self-decomposable deterministic function $f$. The chain decomposition of the model replaces the effect variable $E$ of this partition by $n$ auxiliary variables $E_i$, $i = 1, \ldots, n$, such that

- $E_n$ has the noise variable $Z_n$ for its single parent and encodes the function application $E_n = f(Z_n, I)$, where the variable $I$ captures identity under $f$;
- for all $i = 1, \ldots, n-1$, the variable $E_i$ has $Z_i$ and $E_{i+1}$ for its parents and encodes $E_i = f(Z_i, E_{i+1})$.

If the interaction model includes a leak variable $Z_L$ the identity variable $I$ in the function application $f(Z_n, I)$ is replaced by $Z_L$, to give $E_n = f(Z_n, Z_L)$. We note that the number of variables in the chain decomposition has increased, from $2 \cdot n + 1$ in the original partition, to $3 \cdot n$. The total number of non-redundant probabilities required for the probability tables for the variables $E_i$ in the chain

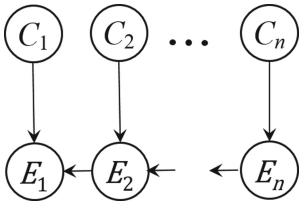



**Fig. 3.** The cascading representation of a causal interaction model, which results from marginalising out the noise variables $Z_i$ from its chain decomposition.

equals $4 \cdot n - 2$ however, instead of the $2^n$ probabilities required for the effect variable $E$ in the original partition. For an interaction model with a leak variable, the number of required probabilities for the effect variable(s) is reduced from $2^{n+1}$ to $4 \cdot n$. We will return to these observations in further detail in Sect. 4.

While the original motivation for partitioning causal interaction models was to underline their induced ease of knowledge acquisition, Heckerman noted that the introduction of the hidden noise variables $Z_i$ in fact made probability elicitation harder rather than easier, as "assessments are easier to elicit (and presumably more reliable) when a person makes them in terms of observable variables" [9]. Following this insight, he proposed a *temporal interpretation* of independence of causal influences for causal interaction models in which a cause $C_i$ is assumed to occur (or not) at time $i$ and has associated its own effect variable $E_i$ indicating the effect after the presence or absence of the first $i$ causes have been observed. With this temporal interpretation, the hidden noise variables are no longer required and the effect variables $E_i$ have in fact become observable variables with a clear semantics supporting probability elicitation. As noted already by Heckerman himself, this temporal interpretation for causal interaction models has reduced applicability for its main drawback [9,10].

## 4 Properties of a Cascading Representation

We propose a representation of causal interaction models that is quite similar to Heckerman's temporal representation, yet without the temporal interpretation. We will argue that our representation has a clear semantics and in addition allows for easy maintenance in the event of changes in the parameters of the represented interaction model. Before demonstrating the latter in Sect. 4.2, we now first detail our *cascading representation* of causal interaction models.

### 4.1 The Cascading Representation and Its Equivalence Property

We focus on causal mechanisms with an underlying self-decomposable deterministic function $f$ as reviewed in the previous section, and consider their chain decomposition as illustrated in Fig. 2 (*right*). Instead of building on a temporal interpretation as suggested by Heckerman, we propose to sum out the noise

variables $Z_i$ by marginalisation. We note that, by doing so, the effect variables $E_i$, $i = 1, \ldots, n$, become stochastic rather than deterministic. The resulting representation, called the *cascading representation* of a causal interaction model, is illustrated for a mechanism $\mathcal{M}(n)$ in Fig. 3, where

– the variable $E_n$, with the cause variable $C_n$ for its single parent, has the probability table derived from the chain decomposition as

$$\Pr(e_n \mid C_n) = \sum_{z'_n \in \Omega(Z_n)} \Pr(e_n \mid z'_n) \cdot \Pr(z'_n \mid C_n) \tag{1}$$

or, in the presence of a leak probability, as

$$\begin{aligned} \Pr(e_n \mid C_n) \quad &= p_L \cdot \sum_{z'_n \in \Omega(Z_n)} \Pr(e_n \mid z'_n, z_L) \cdot \Pr(z'_n \mid C_n) \\ &+ (1 - p_L) \cdot \sum_{z'_n \in \Omega(Z_n)} \Pr(e_n \mid z'_n, \overline{z}_L) \cdot \Pr(z'_n \mid C_n) \end{aligned} \tag{2}$$

– the variables $E_i$, $i = 1, \ldots, n-1$, with the parents $C_i$ and $E_{i+1}$, have the probability table derived as

$$\Pr(e_i \mid C_i, E_{i+1}) \;=\; \sum_{z'_i \in \Omega(Z_i)} \Pr(e_i \mid z'_i, E_{i+1}) \cdot \Pr(z'_i \mid C_i) \tag{3}$$

We note that all probabilities conditioned on a value of a noise variable originate from the degenerate distributions modelling the deterministic function $f$ of the interaction model. We further note that the inclusion of a leak probability affects only the cells of the probability table for the variable $E_n$, whereas it affects, through the definitional rule of the interaction model at hand, *all* cells in the fully expanded table for the variable $E$ in the causal mechanism.

To ensure that our cascading representation of an interaction model is equivalent to its original representation in a causal mechanism, the variable $E_1$ in our representation should represent the exact same information as the effect variable $E$ in a mechanism $\mathcal{M}(n)$. Any probability $\Pr(\overline{e} \mid \mathbf{c}) = 1 - \Pr(e \mid \mathbf{c})$ specified in the full probability table for $E$ should therefore be the same as the probability $\Pr(\overline{e}_1 \mid \mathbf{c})$ that is computed from the cascading representation as

$$\Pr(\overline{e}_1 \mid \mathbf{c}) \;=\; \sum_{\mathbf{e}^- \in \Omega(\mathbf{E}^-)} \Pr(\overline{e}_1 \mid c'_1, e'_2) \cdot \prod_{k=2}^{n-1} \Pr(e'_k \mid c'_k, e'_{k+1}) \cdot \Pr(e'_n \mid c'_n) \tag{4}$$

where $\Omega(\mathbf{E}^-)$ is the domain of the variable set $\mathbf{E}^- = \{E_2, \ldots, E_n\}$, and where $e'_k \in \Omega(E_k)$, $k = 2, \ldots, n$, is consistent with $\mathbf{e}^-$ and $c'_k \in \Omega(C_k)$, $k = 1, \ldots, n$, is consistent with $\mathbf{c}$. We emphasize that we focus on the value $\overline{e}_1$ of the variable $E_1$ rather than on the value $e_1$, to simplify our arguments in the sequel.

We now illustrate the derivation of the probability tables for the cascading representations of the noisy-OR and leaky noisy-OR models, and demonstrate their equivalence to the standard causal-mechanism representation.

*The cascading noisy*-OR. We begin with constructing the conditional probability tables to be specified for the noisy-OR model in its cascading representation. For the variables $E_i$, $i = 1, \ldots, n-1$, we find from Eq. 3 that

$$\begin{aligned}
\Pr(e_i \mid \overline{c}_i, \overline{e}_{i+1}) &= 1 \cdot 0 + 0 \cdot 1 = 0 \\
\Pr(e_i \mid c_i, \overline{e}_{i+1}) &= 1 \cdot p_i + 0 \cdot (1 - p_i) = p_i \\
\Pr(e_i \mid \overline{c}_i, e_{i+1}) &= 1 \cdot 0 + 1 \cdot 1 = 1 \\
\Pr(e_i \mid c_i, e_{i+1}) &= 1 \cdot p_i + 1 \cdot (1 - p_i) = 1
\end{aligned}$$

where $p_i = \Pr(e \mid \overline{c}_1, \ldots, \overline{c}_{i-1}, c_i, \overline{c}_{i+1}, \ldots, \overline{c}_n)$ coincides with a regular noisy-OR parameter. For the variable $E_n$ we similarly find from Eq. 2 that

$$\begin{aligned}
\Pr(e_n \mid \overline{c}_n) &= 1 \cdot 0 + 0 \cdot 1 = 0 \\
\Pr(e_n \mid c_n) &= 1 \cdot p_n + 0 \cdot (1 - p_n) = p_n
\end{aligned}$$

where $p_n$ is again a regular noisy-OR parameter. We observe that each parameter $p_i$, $i = 1, \ldots, n$, occurs in the specification of exactly *one* table, which is the table for the variable $E_i$. In addition to this single associated noisy-OR parameter, the probability table for the variable $E_i$ further specifies just zeroes and ones.

We now show that the cascading representation, with the probability specification above, correctly captures the noisy-OR model. To this end, we observe that for a summand of Eq. 4 to actually contribute to the computation of $\Pr(\overline{e}_1 \mid \mathbf{c})$, it should be a product composed of just non-zero terms. Such non-zero terms are found only with the following probabilities:

- $\Pr(\overline{e}_n \mid \overline{c}_n)$ or $\Pr(e'_n \mid c_n)$, for the variable $E_n$;
- $\Pr(e_i \mid c'_i, e_{i+1})$, $\Pr(\overline{e}_i \mid c'_i, \overline{e}_{i+1})$, and $\Pr(e_i \mid c_i, \overline{e}_{i+1})$, for the variable $E_i$, $i = 1, \ldots, n-1$;

with $e'_i \in \Omega(E_j)$ and $c'_i \in \Omega(C_i)$, $i = 1, \ldots, n$. Close examination of these non-zero probabilities shows that for the value $\overline{e}_1$ of $E_1$ under consideration, only value combinations $\mathbf{e}^-$ for $\mathbf{E}^- = \{E_2, \ldots, E_n\}$ consistent with $\overline{e}_2$ can possibly contribute a non-zero term to a summand of Eq. 4. By iteratively applying this argument to the variables $E_3, \ldots, E_n$, we conclude that only the value combination $\mathbf{e}^- = \overline{e}_2, \ldots, \overline{e}_n$ contributes a non-zero summand to the probability $\Pr(\overline{e}_1 \mid \mathbf{c})$. For the cascading representation of the noisy-OR model therefore, Eq. 4 reduces to:

$$\Pr(\overline{e}_1 \mid \mathbf{c}) = \prod_{i=1}^{n-1} \Pr(\overline{e}_i \mid c'_i, \overline{e}_{i+1}) \cdot \Pr(\overline{e}_n \mid c'_n) \tag{5}$$

To show that the cascading representation correctly captures the noisy-OR model, we now consider the three different cases distinguished by this model:

- Where the noisy-OR model has $\Pr(e \mid \mathbf{c}) = 0$ for $\mathbf{c} = \overline{c}_1, \ldots, \overline{c}_n$, we find in the cascading representation from $\Pr(\overline{e}_j \mid \overline{c}_j, \overline{e}_{j+1}) = 1$ for $j = 1, \ldots, n-1$ and $\Pr(\overline{e}_n \mid \overline{c}_n) = 1$, that $\Pr(\overline{e}_1 \mid \mathbf{c}) = 1$ and, hence, $\Pr(e_1 \mid \mathbf{c}) = 0$.

**Table 1.** For the two representations of the noisy-OR model for a causal mechanism $\mathcal{M}(n)$: the number of variables (*#variables*), the number of non-redundant probabilities for the effect variable(s) (*#probabilities*), and of those, the number of free parameters to be acquired (*#free*) and the number of zeroes and ones (#0/1).

| *Representation* | *#variables* | *#probabilities* | *#free* | #0/1 |
|---|---|---|---|---|
| Full table | $n+1$ | $2^n$ | $n$ | 1 |
| Cascade | $2{\cdot}n$ | $4{\cdot}n-2$ | $n$ | $3{\cdot}n-2$ |

- Where the noisy-OR model has $\Pr(e \mid \mathbf{c}) = p_i$ for $\mathbf{c}$ including the *single* present cause $c_i$, we have in the cascading representation that the product term contributed for the variable $E_i$ has the probability $\Pr(\overline{e}_i \mid c_i, \overline{e}_{i+1}) = 1-p_i$ or, in case $i = n$, $\Pr(\overline{e}_n \mid c_n) = 1-p_n$. As all other terms in the product of Eq. 5 equal 1, we find that $\Pr(\overline{e}_1 \mid \mathbf{c}) = 1 - p_i$ and, hence, $\Pr(e_1 \mid \mathbf{c}) = p_i$.
- For any value combination $\mathbf{c}$ including multiple present causes, with their indices in $I_{\mathbf{c}}$, the noisy-OR model has $\Pr(e \mid \mathbf{c}) = 1 - \prod_{i \in I_{\mathbf{c}}}(1 - p_i)$. In the cascading representation, the product term contributed by any $E_j$ with $j \notin I_{\mathbf{c}}$ equals 1 and the term by any $E_i$ with $i \in I_{\mathbf{c}}$ is $1 - p_i$. We thus find that $\Pr(\overline{e}_1 \mid \mathbf{c}) = \prod_{i \in I_{\mathbf{c}}}(1 - p_i)$ and, hence, $\Pr(e_1 \mid \mathbf{c}) = 1 - \prod_{i \in I_{\mathbf{c}}}(1 - p_i)$.

From the three cases above, we conclude that the cascading representation indeed correctly captures the noisy-OR model and, hence, that the cascading representation is equivalent with the fully expanded probability table for the effect variable $E$ in a causal mechanism with a noisy-OR model.

The cascading representation of the noisy-OR model is a more efficient representation than a causal mechanism $\mathcal{M}(n)$ with a full probability table for the effect variable $E$, despite the increase in number of variables to $2{\cdot}n$ compared to the $n+1$ variables in the standard representation. More specifically, the cascading representation requires $4{\cdot}(n-1)+2$ conditional probability distributions in total for the variables $E_i$, of which $3{\cdot}(n-1)+1$ are degenerate. For ease of reference, Table 1 summarises a comparison of the size of the cascading representation with that of the standard representation. We note that the cascading representation is more concise when a causal mechanism would include $n \geq 4$ cause variables for the effect variable of interest.

*The cascading leaky noisy*-OR. We now briefly address the cascading representation of the noisy-OR model in the presence of a leak probability, which differs from that of the standard noisy-OR model only in the specification of the probability table for the variable $E_n$, which is derived from Eq. 2 as

$$\Pr(e_n \mid \overline{c}_n) = p_L$$
$$\Pr(e_n \mid c_n) = p_L + p_n \cdot (1 - p_L) = 1 - (1 - p_L) \cdot (1 - p_n)$$

where $p_n$ is again a regular noisy-OR parameter and $p_L$ is the leak probability. To show that the cascading representation with this specification correctly

captures the leaky noisy-OR model, we use Eq. 5 again, now for the different cases distinguished by the leaky noisy-OR model. We observe that, while with the noisy-OR model, the variable $E_n$ would contribute to the product either $\Pr(\overline{e}_n \mid \overline{c}_n) = 1$ or $\Pr(\overline{e}_n \mid c_n) = 1 - p_n$, it contributes either $\Pr(\overline{e}_n \mid \overline{c}_n) = 1 - p_L$ or $\Pr(\overline{e}_n \mid c_n) = (1 - p_L) \cdot (1 - p_n)$ in the cascading representation of the leaky noisy-OR model. As a consequence

- With the leaky model having $\Pr(e \mid \mathbf{c}) = p_L$ for $\mathbf{c} = \overline{c}_1, \ldots, \overline{c}_n$, we find $\Pr(\overline{e}_1 \mid \mathbf{c}) = 1 - p_L$ and, hence, $\Pr(e_1 \mid \mathbf{c}) = p_L$ from the cascading representation.
- For any value combination $\mathbf{c}$ with an arbitrary number of present causes with indices in $I_{\mathbf{c}}$, the leaky model has $\Pr(e \mid \mathbf{c}) = 1-(1-p_L)\cdot\prod_{i\in I_{\mathbf{c}}}(1-p_i)$. Using the observation above, we find in the cascading representation that $\Pr(\overline{e}_1 \mid \mathbf{c}) = (1-p_L)\cdot\prod_{j\in I_{\mathbf{c}}}(1-p_j)$ and, hence, $\Pr(e_1 \mid \mathbf{c}) = 1-(1-p_L)\cdot\prod_{j\in I_{\mathbf{c}}}(1-p_j)$.

We conclude that the probabilities computed from the cascading representation indeed coincide with the probabilities in the full probability table in a causal mechanism with the leaky noisy-OR model. We thus can construct an efficient representation for a causal mechanism $\mathcal{M}(n)$ with the leaky noisy-OR model. Of the $4\cdot(n-1)+2$ conditional distributions required in total by the cascading representation, now $3\cdot(n-1)$ are degenerate. We note that the difference of one compared with the cascading representation of the noisy-OR model originates from the inclusion of the leak probability as a parameter.

### 4.2 Additional Engineering Benefits

Causal mechanisms are typically modelled straightforwardly in Bayesian networks, as in Fig. 1 (*left*). The different partitions and decompositions of causal interaction models proposed, are mostly seen as alternative representations to support probability elicitation and are hardly ever used in a network directly. Table 1 clearly illustrates the reduction in specification size that would be achieved by choosing a cascading representation for causal mechanisms with large numbers of cause variables; as this representation limits the number of parents per effect variable, it also has the potential to reduce the runtime complexity of probabilistic inference, dependent of the graphical structure of the embedding Bayesian network [10,14]. In this section, we now argue that the cascading representation further has clear engineering benefits beyond those widely recognised.

*Clear semantics.* Alternative representations of causal interaction models typically rely on the introduction of additional variables. Although introducing such additional variables is commonly used for reducing the number of parents for an effect variable, it is often quite undesirable from a knowledge engineering perspective. While the additional variables have a clear meaning from a mathematics point of view, they often are quite meaningless from the perspective of the application domain and thereby hamper the interpretation of the model as a domain

representation. The lack of a clear meaning is especially problematic if the probabilities for these additional variables need be elicited from experts. Now, in our cascading representation of a causal interaction model, the additional variables *do* have a clear intuitive meaning, as a consequence of the decomposability properties of the underlying deterministic function: in the cascading representation of a causal mechanism $\mathcal{M}(n)$, any variable $E_i$ can be viewed as the effect variable in the causal mechanism $\mathcal{M}(n-i+1)$ involving the subset of causes $C_i, \ldots, C_n$. This claim is readily seen by replacing $E_1$ by $E_i$ in Eq. 4:

$$\Pr(\overline{e}_i \mid \mathbf{c}) = \sum_{\mathbf{e}^- \in \Omega(\mathbf{E}^-)} \Pr(\overline{e}_i \mid c'_1, e'_{i+1}) \cdot \prod_{k=i+1}^{n-1} \Pr(e'_k \mid c'_k, e'_{k+1}) \cdot \Pr(e'_n \mid c'_n)$$

where $\Omega(\mathbf{E}^-)$ now is the domain of $\mathbf{E}^- = \{E_{i+1}, \ldots, E_n\}$, and $e'_k$, $c'_k$ are defined as before. As each variable $E_i$ in the cascading representation represents the effect variable in a (leaky) noisy-OR model with the cause variables $C_i, \ldots, C_n$, it has an intuitive meaning that allows for explicit embedding of the representation in a network without hampering interpretation and probability elicitation.

*Maintenance robustness.* The cascading representation of a causal interaction model brings yet another advantage from an engineering perspective. When using fully expanded probability tables for the effect variables in a Bayesian network, any modelling decision to employ a causal interaction model is no longer explicitly visible in the network's representation. More specifically, the dependency of multiple cells of the table on the parameters of the model employed is hidden. When a network is maintained and adapted to its changing context of application over a period of years therefore, inopportune changes to the specified probabilities can disrupt the modelled interaction pattern and, thereby, the original modelling decision. We illustrate this observation by means of a causal mechanism with a noisy-OR model for the effect variable, and show that the cascading representation of the interaction model used is more robust by preventing the occurrence of such unintended disruptions.

We address the engineering task of studying the effects, on a network's output probabilities, of changing a single probability from one of the network's probability tables. Such a sensitivity analysis is usually part of the encompassing task of fine-tuning the network's specification to attain a desired effect on the output (see for example [1–3]). In view of a causal mechanism $\mathcal{M}(n)$, we now consider the output probability of interest $\Pr(e \,|\, c_i, c_k)$, for some $1 \leq i < k < n$, and address how this probability changes with a change of the probability $x = \Pr(e \mid \overline{c}_1, \ldots, \overline{c}_{i-1}, c_i, \overline{c}_{i+1}, \ldots, \overline{c}_n)$ of the full probability table of the effect variable $E$; we note that this probability is one of the parameters of the noisy-OR model. The function $[\Pr(e \mid c_i, c_k)]\,(x)$ describing the sensitivity of $\Pr(e \mid c_i, c_k)$ to changes in $x$ would be constant if the modelling choice of imposing a noisy-OR interaction for the mechanism at hand is not taken into consideration:

$$[\Pr(e \mid c_i, c_k)](x) = a, \quad \text{with} \quad a = \sum_{\mathbf{c}^- \in \Omega(\mathbf{C}^-)} \Pr(e \mid c_i, c_k, \mathbf{c}^-) \cdot \Pr(\mathbf{c}^- \mid c_i, c_k) \quad (6)$$

where $\Omega(\mathbf{C}^-)$ is the domain of the set $\mathbf{C}^- = \{C_1, \ldots, C_n\} \setminus \{C_i, C_k\}$ of cause variables for which no value is fixed by the probability of interest. We note that the computation of $\Pr(\mathbf{c}^- \mid c_i, c_k)$ does not involve any probabilities from the probability table of $E$; in contrast, the first term in the product for each summand in $a$ corresponds directly to a cell from the full table for $E$. Since the summation does not involve parameter $x$ directly, the analysis reveals that the output probability is not sensitive to variations of the parameter. This result however, does not correctly reflect the true sensitivity of the output probability to variations in the parameter under study: the parameter $x$ is actually included in various cells of the full probability table of $E$ by the definitional rule from the noisy-OR model, and thereby hidden in various summands of $a$.

We now consider the same sensitivity analysis in view of the cascading representation of the noisy-OR model, for essentially the same probability of interest and essentially the same parameter probability. Recall that in the cascading representation, any posterior probability distribution over the variable $E_1$ equals the posterior distribution given the same evidence over the original variable $E$ with the full probability table; we therefore take the probability $\Pr(e_1 \mid c_i, c_k)$ for the probability of interest. The parameter $p_i = \Pr(e \mid \bar{c}_1, \ldots, \bar{c}_{i-1}, c_i, \bar{c}_{i+1}, \ldots, \bar{c}_n)$ of the noisy-OR model moreover occurs as $p_i = \Pr(e_i \mid c_i, \bar{e}_{i+1})$ in the model's cascading representation; we thus take $x = \Pr(e_i \mid c_i, \bar{e}_{i+1})$ as the probability that will be varied. The sensitivity analysis will in essence establish the same result as presented in Eq. 6, but now the probabilities $\Pr(e \mid c_i, c_k, \mathbf{c}^-)$ follow from the cascading representation using Eq. 5, and depend explicitly on $x$:

$$\begin{aligned}
\left[\Pr(\bar{e}_1 \mid c_i, c_k, \mathbf{c}^-)\right](x) &= \left[(1 - p_i) \cdot (1 - p_k) \cdot \prod_{j \in I_{\mathbf{c}^-}} (1 - p_j)\right](x) \\
&= (1 - x) \cdot (1 - p_k) \cdot \prod_{j \in I_{\mathbf{c}^-}} (1 - p_j)
\end{aligned}$$

where $I_{\mathbf{c}^-}$ indexes all present causes in $\mathbf{C}^-$ and, for ease of exposition, we again focus on the value $\bar{e}_1$ for variable $E_1$. As a result, we find that

$$[\Pr(\bar{e}_1 \mid c_i, c_k)](x) = \sum_{\mathbf{c}^- \in \Omega(\mathbf{C}^-)} (1 - x) \cdot (1 - p_k) \cdot \prod_{j \in I_{\mathbf{c}^-}} (1 - p_j) \cdot \Pr(\mathbf{c}^- \mid c_i, c_k)$$

and conclude that the function $[\Pr(e_1 \mid c_i, c_k)](x)$ is in fact a linear function of the form $a \cdot x + b$ with constants $a, b$, where

$$\begin{aligned}
a &= (1 - p_k) \cdot \sum_{\mathbf{c}^- \in \Omega(\mathbf{C}^-)} \Pr(\mathbf{c}^- \mid c_i, c_k) \prod_{j \in I_{\mathbf{c}^-}} (1 - p_j) \\
b &= 1 - a
\end{aligned}$$

The cascading representation of the noisy-OR model performs, during inference, the computation of the probabilities of the effect $e$ given possible combinations of causes. That is, application of the definitional rule is in essence left to inference, resulting in the dependency of the output probability of interest on the noisy-OR parameter now being correctly taken into consideration. This observation further demonstrates that, when changing a single parameter of the noisy-OR model specification upon fine-tuning a Bayesian network, in the cascading representation just a single cell of the conditional probability table for the appropriate effect variable $E_i$ needs to be adapted; in contrast, in the representation with a full conditional probability table, various cells that are specified using the model's definitional rule will need adaptation. The cascading representation is therefore easier to adapt without the risk of violating the properties of the underlying causal interaction model.

## 5 Conclusions and Further Research

In this paper we revisited part of the large volume of work on causal interaction models, and focused thereby on the representational complexity of such models. We built on this early work for the purpose of demonstrating that some of these models allow for a representation with various elegant properties that have not been recognised until now. More specifically, by exploiting the property of self-decomposability of the deterministic function underlying a causal interaction model, we arrived at an alternative cascading representation that has a clear intuitive semantics in terms of the causal mechanism itself, not requiring the inclusion of artificial unobservable variables. In addition to well-known complexity benefits of such alternative representations, this specific cascading representation has important knowledge engineering benefits, allowing easier maintenance and more robust fine-tuning of parameters. As the compactness of the cascading representation can be exploited directly by standard inference algorithms moreover, we conclude all in all that this representation of causal interaction models is quite suitable for explicit embedding in Bayesian networks.

While we used the (leaky) noisy-OR model for our example causal interaction model throughout the paper, the presented properties of the cascading representation apply straightforwardly to any interaction model involving binary-valued variables and having an underlying self-decomposable deterministic function, such as the (leaky) noisy-AND model. For our further research we aim at extending our results to causal interaction models involving multi-valued variables, such as the noisy-MAX model [5], and to other types of decomposable function.

## References

1. Castillo, E., Gutiérrez, J.M., Hadi, A.S.: Sensitivity analysis in discrete Bayesian networks. IEEE Trans. Syst. Man Cybern. **27**, 412–423 (1997)
2. Chan, H., Darwiche, A.: Sensitivity analysis in Bayesian networks: From single to multiple parameters. In: Halpern, J., Meek, C. (eds.) Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, pp. 67–75 (2004)

3. Coupé, V.M.H., van der Gaag, L.C.: Properties of sensitivity analysis of Bayesian belief networks. Ann. Math. Artif. Intell. **36**, 323–356 (2002)
4. Díez, F.J., Druzdzel, M.J.: Canonical Probabilistic Models for Knowledge Engineering. Technical Report CISIAD-06-01 (2007)
5. Díez, F.J., Galán, S.F.: Efficient computation for the noisy max. Int. J. Intell. Syst. **18**, 165–177 (2003)
6. Frey, B.J., Patrascu, R., Jaakkola, T., Moran, J.: Sequentially fitting inclusive trees for inference in noisy- or networks. In: Leen, T.K., Dietterich, T.G., Tresp, V. (eds.) Advances in Neural Information Processing Systems 13, pp. 493–499. MIT Press, Cambridge (2001)
7. Getoor, L.: Learning Statistical Models from Relational Data. PhD Thesis. Stanford University (2001)
8. Heckerman, D.: A tractable inference algorithm for diagnosing multiple diseases. In: Henrion, M., Kanal, L., Lemmer, J., Shachter, R. (eds.) Proceedings of the 5th Conference on Uncertainty in Artificial Intelligence, pp. 163–172 (1989)
9. Heckerman, D.: Causal independence for knowledge acquisition and inference. In: Heckerman, D., Mamdani, E. (eds.) Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence, pp. 122–127 (1993)
10. Heckerman, D., Breese, J.: Causal independence for probability assessment and inference using Bayesian networks. IEEE Trans. Syst. Man Cybern. **26**, 826–831 (1996)
11. Henrion, M.: Some practical issues in constructing belief networks. In: Kanal, L.N., Levitt, T.S., Lemmer, J.F. (eds.) Uncertainty in Artificial Intelligence 3, pp. 161–173. Elsevier (1989)
12. Huang, K., Henrion, M.: Efficient search-based inference for noisy- or belief networks: TopEpsilon. In: Horvitz, E., Jensen, F. (eds.) Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence, pp. 325–331 (1996)
13. Jesus, P., Baquero, C., Almeida, P.S.: A survey of distributed data aggregation algorithms. IEEE Commun. Surv. Tutorials **17**, 381–404 (2011)
14. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. The MIT Press, Cambridge (2009)
15. Li, W., Poupart, P., van Beek, P.: Exploiting structure in weighted model counting approaches to probabilistic inference. J. Artif. Intell. Res. **40**, 729–765 (2011)
16. Olesen, K.G., et al.: A MUNIN network for the median nerve: a case study on loops. Appl. Artif. Intell. **3**, 385–403 (1989)
17. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, Burlington (1988)
18. del Sagrado, J., Salmerón, A.: Representing canonical models as probability trees. In: Conejo, R., Urretavizcaya, M., Pérez-de-la-Cruz, J.-L. (eds.) CAEPIA/TTIA -2003. LNCS (LNAI), vol. 3040, pp. 478–487. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-25945-9_47
19. Zhang, N.L., Yan, L.: Independence of causal influence and clique tree propagation. Int. J. Approximate Reasoning **19**, 335–349 (1998)