Recursive Noisy OR—A Rule for Estimating Complex Probabilistic Interactions

John F. Lemmer and Don E. Gossink

***Abstract*—This paper focuses on approaches that address the intractability of knowledge acquisition of conditional probability tables in causal or Bayesian belief networks. We state a rule that we term the "recursive noisy OR" (RNOR) which allows combinations of dependent causes to be entered and later used for estimating the probability of an effect. In the development of this paper, we investigate the axiomatic correctness and semantic meaning of this rule and show that the recursive noisy OR is a generalization of the well-known noisy OR. We introduce the concept of positive causality and demonstrate its utility in axiomatic correctness of the RNOR. We also introduce concepts describing the ways in which dependent causes can work together as being either "synergistic" or "interfering." We provide a formalization to quantify these concepts and show that they are preserved by the RNOR. Finally, we present a method for the determination of Conditional Probability Tables from this causal theory.**

***Index Terms*—Bayes, causality, estimation, uncertainty.**

I. Introduction

In this paper, a theory allowing the tractable construction of more accurate conditional probability tables (CPT) for Bayesian causal models is presented. The potentially huge effort associated with construction of accurate CPTs is sometimes referred to as "the intractability of knowledge acquisition" [6] or "knowledge acquisition bottleneck" [7] problem. Predominantly, and for this paper, the allowable states for the variables are binary valued.[1]

Tractability of CPT construction has risen as a practical issue in construction of real models in which CPTs represent expert opinion and absence of data precludes learning approaches. It is not uncommon for an effect to have eight to ten and sometimes more relevant causes which translates to CPTs requiring from 256 to 1024 entries. Model builders simply will not expend the effort required to estimate this many parameters.

Accuracy of CPTs has been limited by the current techniques used for actually estimating CPTs from limited user inputs. The primary technique in use today is the Noisy OR model attributed to Pearl [4] but traceable to Good [17], [18], and extended by others [7], [9], [10]. The Noisy OR theory assumes that all causes of an effect are independent of each other: that each cause is capable, by itself, of causing the effect, and that this capability is not affected by the presence or absence of other active causes. Thus, using Noisy OR, it is not possible to capture such causal notions as synergy among causes, interference, necessity, and the like. Yet such notions often dominate the thinking of actual model builders.

The theory to be presented goes beyond probability theory by using causal ideas to make CPT construction more tractable and accurate. Strict probability theory provides no bounds on the entries in a CPT, except each entry must be a value within the range zero to one. The Noisy OR formalism mentioned in Pearl [4] can be viewed as an element of a causal theory, even though it is usually presented as a probabilistic theory. In fact the work presented here is essentially an extension of Noisy OR The theory to be developed here relates various CPT values to each other, given various causal assumptions. To emphasize the causal nature of Noisy OR, we will develop it from a causal foundation rather than from the usual probabilistic-logical foundations. This development is presented in Section II of the paper.

After developing the Noisy OR from a causal foundation, in Section III we will extend Noisy OR to a technique we term Recursive Noisy OR (RNOR). RNOR allows arbitrary entries of causally dependent probabilities to be entered. These arbitrary entries are intended to be the modelers' key estimates about a domain's synergy, AND relations, etc, the very estimates which cannot be captured by Noisy OR itself. Most importantly the rule allows the causal information in the provided values to be used in a semantically consistent way to estimate other entries which the modeler chooses not to provide. The rule is analyzed and its limitations are discussed. We also present an algorithm using RNOR to construct a complete CPT. We claim that this algorithm allows domain causal relationships, over and above causal independence to be tractably captured in a semantically meaningful way. With this algorithm, more accurate Bayes models [1]–[5] can tractably be built. In our experience, we have seen models in which this increased accuracy and tractability has rendered the Bayesian approach usable.

II. Noisy OR: A Model of Independent Cause in Uncertainty

Pearl [4], [8] is credited with the introduction of the Noisy OR for binary variables, however it is observable in the work of Good [17], [18]. This was done under a so-called modeling type of "disjunctive interaction" leading to a widely applicable rule of combination. For the original derivation of the Noisy OR the reader is referred to Pearl [4]. As mentioned previously, we will develop the Noisy OR from a causal foundation.

An uncertain cause is a cause which *can* produce an effect, but does not always do so. Throwing a ball at bowling pin supported

Manuscript received October 15, 2003; revised March 11, 2004. This paper was recommended by Associate Editor E. Santos.

J. F. Lemmer is with the Air Force Research Laboratory, Information Directorate, Rome, NY 13441-4514 USA (e-mail: John.Lemmer@rl.af.mil).

D. E. Gossink is with the Command and Control Division, Defence Science and Technology Organization, Edinburgh 5111, Australia (e-mail: Don.Gossink@dsto.defence.gov.au).

Digital Object Identifier 10.1109/TSMCB.2004.834424

[1]Allowable states are $\mathrm{False} = 0$ and $\mathrm{True} = 1$.

on a table is such an uncertain cause: the ball may not strike the pin, but even if it does, the pin may not fall over. Thus, we can talk of the probability that a ball thrown by some individual will knock over a pin.

An independent cause is one whose probability of effect is not dependent on any other causes. Multiple throws of a ball at a bowling pin can be such independent causes. If any particular throw has a 30% chance of knocking over the pin, it makes no difference if three or if five previous throws have failed to knock the pin, the probability on the next throw remains 30%. Independent causes do not need to be repetitions of the same cause. For example, kicking the table on which the pin is standing could be another independent cause for knocking the pin over.

Under the assumption of independent causality, all causes must have each failed (independently) to achieve its purpose if the effect has not been achieved. If the pin remains standing, then all throws must have failed to knock the pin over and all kicks failed as well. If there have been two throws and one kick, we can use independence of the causes to compute that the probability of the pin not being toppled by one of these causal actions is given by

$$\begin{aligned} p(\text{pin still standing}) &= p(\text{throw 1 didn't do it}) \\ &\quad \times p(\text{throw 2 didn't do it}) \times p(\text{kick didn't do it}). \end{aligned}$$

Given our assumptions, this is equivalent to writing

$$\begin{aligned} (1-p(\text{pin fell over})) &= (1-p(\text{throw 1 knocked it over})) \\ &\quad \times (1-p(\text{throw 2 knocked it over})) \\ &\quad \times (1-p(\text{kick knocked it over})) \end{aligned}$$

or equivalently

$$p(\text{pin fell over}) = \begin{pmatrix} 1-(1-p(\text{throw 1 knocked it over})) \\ \times (1-p(\text{throw 2 knocked it over})) \\ \times (1-p(\text{kick knocked it over})) \end{pmatrix}.$$

More precisely, p (pin fell over) is an estimate of the lower bound of the probability that the pin fell over. Only under the assumption there are no other active causes can we rigorously state that it is the actual probability.

By formalizing and generalizing the above equations, we will derive a rule equivalent to the Noisy OR rule of Pearl [4] (and others). However, we have derived it using only (except for probabilistic independence) causal notions. Formally, we write $p(y \leftarrow c(x))$ to mean the probability that y is caused by x. We use $c(\cdot)$ to emphasize that this is a domain notion of causality; that is, we mean that y is an uncertain outcome of physical process of x in some domain. We use $\leftarrow$ rather than than $|$ to emphasize that our notation does not necessarily represent a conditional probability. Remember the adage that "Correlation is not causation." We will show the relation between $p(y \leftarrow c(x))$ and $p(y|x)$ later in developing our algorithm using RNOR to build a complete CPT.

We can generalize our previous discussion of the probability that the pin fell over by representing that there may be n such independent causes of y, where we use x_i to indicate the ith independent cause. As such, we can now write the general form

$$p\left(y \leftarrow c(x_0, x_1, \ldots, x_{n-1})\right) = 1 - \prod_{i=0}^{n-1} \left(1 - p\left(y \leftarrow c(x_i)\right)\right) \tag{1}$$

meaning that, under the assumptions of independent causality, the probability that y will be caused by at least one of the x's is computed by the expression on the right of (1). Indeed, from a causal theory point of view, (1) can be viewed as our *definition* of independent causality. Note that this equation makes empirically testable predictions.

The form of the equation that will be used most frequently in what follows is

$$\left(1-p\left(y \leftarrow c(x_0, x_1, \ldots, x_{n-1})\right)\right) = \prod_{i=0}^{n-1} \left(1-p\left(y \leftarrow c(x_i)\right)\right). \tag{2}$$

For convenience, we introduce the following shorthand notation: $p(y \leftarrow c(x_0, x_1, \ldots, x_{n-1})) \triangleq \vec{p}_y(x_0, x_1, \ldots x_{n-1})$, which if $\mathbf{x} = \{x_0, x_1, \ldots, x_{n-1}\}$ can be written as $\vec{p}_y(\mathbf{x})$. This can be read as the probability of y caused by $\mathbf{x}$, where $\mathbf{x}$ is the set of active causes of y.[2] We can now rewrite (2) as

$$\left(1-\vec{p}_y(x_0, x_1, \ldots, x_{n-1})\right) = \left(1-\vec{p}_y^{\,N}(\mathbf{x})s\right) = \prod_{x \in \mathbf{x}^*} \left(1-\vec{p}_y(x)\right). \tag{3}$$

We will use a superscript of N to clarify when we are referring to a probability estimate generated by the Noisy OR.

We will show in Section III-B how the results of this and the next section can be used to create CPTs in the context of a Bayesian network. But nowhere yet have we relied on Bayesian Network concepts. Likewise, the probabilities we have introduced so far are not "conditional" probabilities in the probabilistic sense. They are probabilities (potentially at least) determined subjectively. Probabilities such as these which are not conditional probabilities will continue to be the basis of the work in Section III. To relate (2) and (3) to other work in the literature, the probabilities of (2) can be *roughly thought of* as conditional probabilities derived from some joint distribution, and causes and effects thought of as binary events in this joint distribution (e.g. cause "active" is event "true").

A. Related Work

The Noisy OR has also been extended to consider nonbinary (multistate) cases which are termed the Generalized Noisy OR (GNOR) or noisy-MAX. Henrion [10] suggested that "one can treat an n-ary variable as $n-1$ binary variables. The resulting value of the effect variable is the maximum of the levels produced by each of its influencing variables". This approach was extended by Diez [7]. Independently Srinivas [9] derived a version of the GNOR extending the Noisy OR [4] to manage multistate instances through a function which maps the interaction of the multistate inputs to a multistate output. For some instances such as a known digital circuit the determination of this function is a clear-cut exercise, however there are instances where determination of this function is not as evident. In all cases of

[2] $\mathbf{x}$ is the subset of parents in the "true" state.

a GNOR, it has been shown that the original Noisy OR is subsumed.

III. Recursive Noisy OR: Allowing Dependent Causes

In this section, we state a rule allowing combinations of *dependent* causes to be used for estimating the probability of an effect. This rule, as derived in Appendix A, is based on a re-representation of the rule in (3). This representation allows probabilistic knowledge about dependent causal interactions to be used for estimation of the probability of an effect.

As an example of dependent cause, we expand the bowling pin example of the previous section: If we kick the table supporting the bowling pin while the pin is still wobbling from being hit by a throw, there will be a much better chance of the pin falling. It will be better than if we wait for the pin to stop wobbling before kicking. When the throw and the kick occur more or less simultaneously to produce a common effect, they act synergistically. In the synergistic case, the throw and the kick are causally dependent because we feel that (3) is no longer a good approximation of our knowledge.

Let $\vec{p}_f(t)$ be the probability of the bowling pin falling from just a throw, and let $\vec{p}_f(k)$ be the probability from just a kick. By saying that the throw and the kick acting together are not independent, we mean that $(1-\vec{p}_f(t,k)) \neq (1-\vec{p}_f(t))(1-\vec{p}_f(k))$. We can suppose, for example, that we have some empirical (or thought-to-be) values for these probabilities and that the inequality holds. Suppose T is another throw, not related to the throw t or the kick k. We might reasonably compute $\vec{p}_f(T,t,k)$, the probability of the pin falling by three of these causes (including two dependent causes) as $\vec{p}_f(T,t,k) = \vec{p}_f(T)\vec{p}_f(t,k)$. But if there are other dependencies involved (perhaps, the thrower is more likely to succeed on a second throw), a reasonable formula for all three causes is less clear.

To handle the general case of causal dependencies, we propose the following rule, which we term the RNOR (a methodology for deriving the lower part of the representation for the RNOR (4) is presented in Appendix A):

$$p^R(\mathbf{x}) = \begin{cases} p^E(\mathbf{x}), \\ \quad \text{If the expert/modeler provides} \\ \quad \text{this information} \\ 1 - \prod\limits_{i=o}^{n-1} \frac{\left(1-p^R(\mathbf{x}\setminus\{x_i\})\right)}{\left(1-p^R\left(\mathbf{x}\setminus\left\{x_i, x_{(i+1) \bmod n}\right\}\right)\right)}, \\ \quad \text{Otherwise} \end{cases} \quad (4)$$

The symbol $\setminus$ denotes set subtraction and (a) mod (b) represents the modulus function.

For the purposes of readability, we have elected to simplify our notation; instead of explicitly writing $\vec{p}_y(\cdot)$ everywhere, we will use the form $p(\cdot)$ with the inference being that we are discussing causal probabilities as laid out in Section II. Also, for purposes of clarity, we will denote a causal probability of an effect for a set of causes $\mathbf{x}$ to have one of the following superscripts E: $p^E(\mathbf{x})$, N: $p^N(\mathbf{x})$ and R: $p^R(\mathbf{x})$; these are causal probability estimates provided by an Expert, Noisy OR, and RNOR respectively. As with the Noisy OR, absence of active causes in the RNOR induces a zero probability i.e. $p^R(\emptyset) = p^N(\emptyset) = 0$. The RNOR is recursive in the same sense that the factorial function is recursive, that is unknown right hand elements of (4) invoke the RNOR for a smaller set of causes.

The major advantage of the RNOR rule (4) is that it allows arbitrary causally dependent subsets of probability information to be incorporated in the estimation of $p^R(\mathbf{x})$. Thus this rule can handle cases such as described in the earlier paragraph of the bowling pin example. Example 1 below demonstrates the workings of this rule. The ground instance of this recursion occurs when $|\boldsymbol{x}| = 2$. As an example of grounding, note that if $\mathbf{x} = \{a,b\}$, then $1 - p^R(\mathbf{x}) = (1-p^E(a))(1-p^E(b))$. Thus, for (4) to be useful, the probabilities for singleton causes must be provided.

Example 1: The following is an example of how to use the RNOR. Assume that $\mathbf{x} = \{a,b,c\}$. (Then in this case $\mathbf{x}\setminus\{c\} = \{a,b\}$). Thus, (4) means that

$$\left(1-\vec{p}^R(a,b,c)\right) = \frac{\left(1-\vec{p}^R(a,b)\right)\left(1-\vec{p}^R(a,c)\right)\left(1-\vec{p}^R(b,c)\right)}{\left(1-\vec{p}^R(a)\right)\left(1-\vec{p}^R(b)\right)\left(1-\vec{p}^R(c)\right)}. \quad (5)$$

From this example, we can see that the $p^R(\mathbf{x} \setminus \{x_i\})$ and $p^R(\mathbf{x}\setminus\{x_i, x_{(i+1) \bmod n}\})$ appearing in the right part of (4) represents *potentially* causally dependent probabilities. Continuing, if we have an empirical estimate of $p(a,b)$ such that causal independence does not in fact hold, we can substitute this estimate into (5). The new result will be a different estimate of $(1-p(a,b,c))$ than the simple Noisy OR of (3). In the sense that it now uses our knowledge of causal dependence, it may be a better estimate as well as a different estimate.

We base our claims that RNOR is a good rule on several facts. First, RNOR is a generalization of Noisy OR. We will prove that RNOR reduces to Noisy OR when the expert provides $p^E(\mathbf{x})$ only in cases where $|\boldsymbol{x}| = 1$. Second, we will show that certain ratios that we will term synergies and interferences are preserved by this rule. Finally, we will present examples of probabilities estimated by this rule and discuss the appropriateness of these estimates.

We also discuss the conditions under which RNOR can be applied. We will show that as long as the expert provides values which do not imply inhibition among causes, RNOR can be applicable. However, expert provided values implying inhibition will cause RNOR to produce values greater than one. We use a different approach than RNOR to handle inhibition and is discussed in a forthcoming paper.

Before discussing the suitability and usefulness of (4) as a rule into which known (elicited) dependent probability information can be inserted, combined, and utilized, we present a detailed example.

Example 2: For the case where we have $n = |\mathbf{x}| = 4$ causes $(\mathbf{x} = \{x_0, x_1, x_2, x_3\})$ of y. The information that is given is $p^E(x_0) = 0.2$, $p^E(x_1) = 0.25$, $p^E(x_2) = 0.3$, $p^E(x_3) = 0.35$, $p^E(x_0,x_3) = 0.57$, $p^E(x_2,x_3) = 0.68$, $p^E(x_0,x_1,x_2) = 0.75$, and $p^E(x_0,x_2,x_3) = 0.73$. Utilizing the given information, we will compute the probability estimates as per the RNOR and Noisy OR (for comparison)—this is shown in Table I.

The table demonstrates how one uses the RNOR and how either previously computed values or given information is substituted into the RNOR. One should note that given information is

TABLE I
COMPUTATION OF CAUSAL PROBABILITIES, $\vec{p}_y(\cdot)$

Denoted	Recursive OR			Noisy OR
	Calculation	Substitution	Result	
$\vec{p}^{\cdot}(x_0)$	Given		0.2	0.2
$\vec{p}^{\cdot}(x_1)$	Given		0.25	0.25
$\vec{p}^{\cdot}(x_2)$	Given		0.3	0.3
$\vec{p}^{\cdot}(x_3)$	Given		0.35	0.35
$\vec{p}^{\cdot}(x_0,x_1)$	$\left(1-\left(1-\vec{p}^R(x_0)\right)\left(1-\vec{p}^R(x_1)\right)\right)$	$\left(1-(0.8\times 0.75)\right)$	0.4	0.4
$\vec{p}^{\cdot}(x_0,x_2)$	$\left(1-\left(1-\vec{p}^R(x_0)\right)\left(1-\vec{p}^R(x_2)\right)\right)$	$\left(1-(0.8\times 0.7)\right)$	0.44	0.44
$\vec{p}^{\cdot}(x_0,x_3)$	Given		0.57	0.48
$\vec{p}^{\cdot}(x_1,x_2)$	$\left(1-\left(1-\vec{p}^R(x_1)\right)\left(1-\vec{p}^R(x_2)\right)\right)$	$\left(1-(0.75\times 0.7)\right)$	0.475	0.475
$\vec{p}^{\cdot}(x_1,x_3)$	$\left(1-\left(1-\vec{p}_y^R(x_1)\right)\left(1-\vec{p}_y^R(x_3)\right)\right)$	$\left(1-(0.75\times 0.65)\right)$	0.5125	0.5125
$\vec{p}^{\cdot}(x_2,x_3)$	Given		0.68	0.545
$\vec{p}^{\cdot}(x_0,x_1,x_2)$	Given		0.75	0.58
$\vec{p}^{\cdot}(x_0,x_1,x_3)$	$\left(1-\left(\frac{1-\vec{p}^R(x_1,x_3)}{1-\vec{p}^R(x_3)}\times\frac{1-\vec{p}^R(x_0,x_3)}{1-\vec{p}^R(x_0)}\times\frac{1-\vec{p}^R(x_0,x_1)}{1-\vec{p}^R(x_1)}\right)\right)$	$\left(1-\left(\frac{0.4875}{0.65}\frac{0.43}{0.8}\frac{0.6}{0.75}\right)\right)$	0.6775	0.61
$\vec{p}^{\cdot}(x_0,x_2,x_3)$	Given		0.73	0.636
$\vec{p}^{\cdot}(x_1,x_2,x_3)$	$\left(1-\left(\frac{1-\vec{p}^R(x_2,x_3)}{1-\vec{p}^R(x_3)}\times\frac{1-\vec{p}^R(x_1,x_3)}{1-\vec{p}^R(x_1)}\times\frac{1-\vec{p}^R(x_1,x_2)}{1-\vec{p}^R(x_2)}\right)\right)$	$\left(1-\left(\frac{0.32}{0.65}\frac{0.4875}{0.75}\frac{0.525}{0.7}\right)\right)$	0.76	0.65875
$\vec{p}^{\cdot}(x_0,x_1,x_2,x_3)$	$\left(1-\left(\frac{1-\vec{p}^R(x_1,x_2,x_3)}{1-\vec{p}^R(x_2,x_3)}\times\frac{1-\vec{p}^R(x_0,x_2,x_3)}{1-\vec{p}^R(x_0,x_3)}\times\frac{1-\vec{p}^R(x_0,x_1,x_3)}{1-\vec{p}^R(x_0,x_1)}\times\frac{1-\vec{p}^R(x_0,x_1,x_2)}{1-\vec{p}^R(x_1,x_2)}\right)\right)$	$\left(1-\left(\frac{1-0.76}{1-0.68}\times\frac{1-0.73}{1-0.57}\times\frac{1-0.6775}{1-0.4}\times\frac{1-0.75}{1-0.475}\right)\right)$	0.879	0.727

always substituted in preference to a computed value. The example values used in computation of the values of Table I are such that the causal dependencies are moderately synergistic (we will explain our interpretation of synergy later). That is, compared to values generated by the Noisy OR: $p^R(x_0, x_3)$ is 19% greater, $p^R(x_2, x_3)$ is 25% greater, $p^R(x_0, x_1, x_2)$ is 29% greater, and $p^R(x_0, x_2, x_3)$ is 14% greater. Given this information it is reasonable to expect that other combinations would also exhibit a synergistic outcome. The RNOR is such a rule of combination that it allows us to compute an estimate of the causal probability values. As anticipated, the values computed by the RNOR from the information provided do retain a synergistic factor over that of the Noisy OR. For example $p^R(x_0, x_1, x_2, x_3)$ computed via the RNOR is 21% greater than the Noisy OR value.

A. Analysis of the RNOR Rule

There are at least two criteria by which we ought to judge RNOR. The first is: whether the numbers computed by the rule are semantically meaningful. The second is: under what conditions is it axiomatically correct. We could adopt the attitude (prevalent in Fuzzy Set and Possibility Theory communities) to not worry about empirical semantics. However we feel that since RNOR is a causal theory, it requires empirical semantic justification. We are unable to provide *formal* causal justification, but will provide both suggestive theorems and suggestive examples. We will also prove that under a wide range of precisely characterized and useful circumstances, the rule produces axiomatically correct results.

1) Positive Causality: The condition that we term positive causality will be sufficient to ensure axiomatic correctness of probabilities computed by rule (4). The basic idea of positive causality is that additional active causes always increase the probability of achieving an effect.

More formally, we state the following.

Definition 1: Positive Causality

$$\forall \mathbf{z} \subset \mathbf{x} \left\lfloor p^R(\mathbf{z}) \leq p^R(\mathbf{x}) \right\rfloor$$

Intuitively, definition one means that additional causes never decrease the probability of an effect. Examples of this definition follow the next theorem and its proof. The condition of positive causality is sufficient to guarantee that if all expert-provided probabilities $p^E(\cdot)$ are valid probabilities which satisfy positive causality, then all probabilities computed by RNOR are also valid and in general satisfy positive causality.

Theorem 1: If, for all $\mathbf{z}$ such that $\mathbf{z}$ is a proper subset of $\mathbf{x}$, (i.e. $[(\mathbf{z} \subset \mathbf{x}) \wedge (\mathbf{z} \neq \mathbf{x})]$), and $p^R(\mathbf{z})$ is valid:

$$0 \leq p^R(\mathbf{z}) \leq 1.0$$

and positive causality holds

$$p^R(\mathbf{z}) \leq p^R(\mathbf{x})$$

then $p^R(\mathbf{x})$ is valid

$$0 \leq p^R(\mathbf{x}) \leq 1.0.$$

Proof: Recalling that $p^R(\emptyset)=0$

If the expert provides $p^E(\mathbf{x})$, then $p^R(\mathbf{x}) = p^E(\mathbf{x})$ and the theorem is true if $p^E(\mathbf{x})$ is valid.

Otherwise, if $p^R(\mathbf{x})$ is to be computed, then

$$\left(1-p^R(\mathbf{x})\right) = \prod_{i=0}^{n-1} \frac{\left(1-p^R\left(\mathbf{x}\setminus\{x_i\}\right)\right)}{\left(1-p^R\left(\mathbf{x}\setminus\left\{x_i, x_{(i+1)\bmod n}\right\}\right)\right)}.$$

By positive causality, numerator and denominator terms can be paired so that each pair is in the range zero through one. Specifically

$$\forall i, (0\le i<n)\left[0\le \frac{\left(1-p^R(\mathbf{x}\setminus\{x_i\})\right)}{\left(1-p^R\left(\mathbf{x}\setminus\left\{x_i, x_{(i+1)\bmod n}\right\}\right)\right)}\le 1.0\right].$$

From this, we can see that $(1-p^R(\mathbf{x}))$ is a product of terms all of which lie in the inclusive range, 0 through 1. Thus $(1-p^R(\mathbf{x}))$ also lies in this range, which in turn implies that $p^R(\mathbf{x})$ lies in the same range. Thus, $0\le p^R(\mathbf{x})\le 1$. Q.E.D.

An example of positive causality in action is where we have three possible causes of y, $\mathbf{x}=\{a,b,c\}$, we recognize that there are three possibilities for $\mathbf{z}$. That is $\mathbf{z}$ can be either $\{a,b\}$, $\{a,c\}$ or $\{b,c\}$. Under positive causality, this means that $p(a,b,c)\ge \max(p(a,b),p(a,c),p(b,c))$.

Violation of definition one (positive causality) is in fact inhibition, a subject of a different paper. Checking for such violations is trivial and may be considered an appropriate mechanism for model data scrutiny. Positive causality guarantees that RNOR produces estimates within the correct ranges of probability. That is the impact of positive causality is such that the resultant denominator of (4) is always greater than the resultant numerator of (4), thus guaranteeing an estimate within the range of zero and one. Positive causality is sufficient for validity of computed $p^R(\mathbf{x})$'s but is by no means necessary. To see this, consider the proof of the last theorem. It is not necessary that *each and every one* of the terms, t, be in the range $0\le t\le 1$ as guaranteed by positive causality, but only that the full product be in this range.

Unfortunately, positive causality does not guarantee continuing positive causality either. Even if positive causality holds for $\mathbf{x}$, it may not hold for $p^R(\mathbf{x}\cup\{w\})$. To see this consider the following example: Suppose that

$$p^E(a)=p^E(a,b)=p^E(b)=p^E(b,c)=p^E(c)=p^E(a,c)=v$$

where v is some valid probability value. This set of values satisfies positive causality. However, $p^R(a,b,c) = 0$ (why 0 is a reasonable result will be discussed below.) Thus, if this expanded sequence were used to compute, for example, $p^R(a,b,c,d)$, positive causality may also no longer hold. But we should remember that positive causality, while sufficient, is not necessary for the computed $p^R(a,b,c,d)$ to be valid.

Why is it reasonable that $p^R(a,b,c)=0$? Consider the semantics of the values which produced this result. The equality, $p^E(a) = p^E(a,b)$, is semantically interpretable as b has no causal effect when combined with a. Similarly c has no effect combined with b, and a none with c. The net result is that they cancel each other out and resulting causal probability is zero. To observe the zero effect resulting in $p^R(a,b,c)$ substitute the value of v for $p^E(a), p^E(b), p^E(c), p^E(a,b), p^E(b,c), p^E(a,c)$ into the RNOR

$$\begin{aligned} p^R(a,b,c) &= 1-\frac{\left(1-p^R(b,c)\right)\left(1-p^R(a,c)\right)\left(1-p^R(a,b)\right)}{(1-p^R(c))(1-p^R(a))(1-p^R(b))}\\ &= 1-\frac{(1-v)(1-v)(1-v)}{(1-v)(1-v)(1-v)}\\ &= 1-1=0 \end{aligned}$$

A potential problem with RNOR is that its computation can become undefined. This will happen if any $p^R(\cdot)$ appearing in the denominator of the RNOR is equal to one. This is simply ameliorated by noting that if no $p^E(\cdot)$ provided by an expert is ever equal to one, no $p^R(\cdot)$ will become equal to one either.

Theorem 2: If

$$\forall p^E(\cdot)\left\lfloor p^E(\cdot)\ne 1\right\rfloor$$

then

$$\forall \mathbf{z}\subset\mathbf{x}\left\lfloor p^R(\mathbf{z})<1\right\rfloor.$$

Proof: Induction on $|\mathbf{x}|$

Case $|\mathbf{x}|=1$

Assume $\mathbf{x}=\{a\}$. Then $p^R(a)=p^E(a)$ which by hypothesis is not equal to 1.

Case $|\mathbf{x}|>1$

By RNOR

$$p^R(\mathbf{x}) = 1-\prod_{i=0}^{n-1} \frac{\left(1-p^R\left(\mathbf{x}\setminus\{x_i\}\right)\right)}{\left(1-p^R\left(\mathbf{x}\setminus\left\{x_i, x_{(i+1)\bmod n}\right\}\right)\right)}$$

By induction, the numerator will never be 0. Therefore, $p^R(\mathbf{x})$ can never be 1. Q.E.D.

2) Noisy OR *is a Special Case of RNOR:* Here, we will now show that when the expert provides only values, $p^E(\mathbf{x})$, such that $|\boldsymbol{x}|=1$, RNOR is equivalent to Noisy OR.

We have defined the following three ways of estimating $p(\mathbf{x})$.

I. The model builder provides an estimate of $p(\mathbf{x})$. We denote such a probability as $p^E(\mathbf{x})$,

II. Utilizing the Noisy OR as per the literature: if $0<|\mathbf{x}|=n$ then $p^N(\mathbf{x})=1-\prod_{i=0}^{n-1}(1-p^E(x_i))$.

III. The RNOR to be denoted $p^R(\mathbf{x})$, for $|\mathbf{x}|>0$.

$$p^R(\mathbf{x}) = \begin{cases} p^E(\mathbf{x}), \\ \quad \text{If the expert/modeler provides} \\ \quad \text{this information} \\ 1-\prod_{i=o}^{n-1}\frac{\left(1-p^R(\mathbf{x}\setminus\{x_i\})\right)}{\left(1-p^R\left(\mathbf{x}\setminus\left\{x_i,x_{(i+1)}\bmod n\right\}\right)\right)} \\ \quad \text{Otherwise}\end{cases}$$

Note that neither Noisy OR nor RNOR is defined if the modeler does not provide probabilities, $p^E(x_i)$, for all the elements of $\mathbf{x}$. However, these values may be 0.

We now state and prove two lemmas which will help in proving our central results.

Lemma 1: For $|\mathbf{x}|=1$, by definition

$$p^R(\mathbf{x}) = p^N(\mathbf{x}) = p^E(\mathbf{x}).$$

Lemma 2: If $|\mathbf{x}| > 1$, then

$$\prod_{i=0}^{n-1} \frac{\left(1 - p^N\left(\mathbf{x} \setminus \{x_i\}\right)\right)}{\left(1 - p^N\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)\right)} = \left(1 - p^N(\mathbf{x})\right).$$

Proof: We commence with the right hand side of Lemma 2

$$\prod_{i=0}^{n-1} \frac{\left(1 - p^N\left(\mathbf{x} \setminus \{x_i\}\right)\right)}{\left(1 - p^N\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)\right)}.$$

By definition of Noisy OR (3)

$$= \prod_{i=0}^{n-1} \frac{\prod\limits_{x \in (\mathbf{x} \setminus \{x_i\})} \left(1 - p^N(x)\right)}{\prod\limits_{x \in \left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)} \left(1 - p^N(x)\right)}.$$

Because only one term from the numerator is missing in the denominator

$$= \prod_{i=0}^{n-1} \left(1 - p^N\left(x_{(i+1)\bmod n}\right)\right).$$

By definition of the "mod" operator

$$= \prod_{i=0}^{n-1} \left(1 - p^N(x_i)\right)$$

By definition of Noisy OR (3)

$$= \left(1 - p^N(\mathbf{x})\right) \qquad \text{Q.E.D.}$$

The following theorem states that if the only information provided by an expert or modeler is that involving the individual causes, RNOR is equivalent to the Noisy OR.

Theorem 3: If for all $\mathbf{x}$, $|\mathbf{x}| > 1$, $p^E(\mathbf{x})$ is not provided, then

$$p^R(\mathbf{x}) = p^N(\mathbf{x}).$$

Proof: By induction on $|\mathbf{x}|$.

Case $|\mathbf{x}| = 1$:

Let $\mathbf{x} = \{a\}$, Then by Lemma 1

$$\left(1 - p^R(a)\right) = \left(1 - p^N(a)\right) = \left(1 - p^E(a)\right).$$

Case $|\mathbf{x}| > 1$:

Induction hypothesis: The theorem holds for $|\mathbf{x}| - 1$.

By definition

$$\left(1 - p^R(\mathbf{x})\right) = \prod_{i=0}^{n-1} \frac{\left(1 - p^R\left(\mathbf{x} \setminus \{x_i\}\right)\right)}{\left(1 - p^R\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)\right)}.$$

By induction

$$= \prod_{i=0}^{n-1} \frac{\left(1 - p^N\left(\mathbf{x} \setminus \{x_i\}\right)\right)}{\left(1 - p^N\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)\right)}.$$

By Lemma 2

$$= \left(1 - p^N(\mathbf{x})\right). \qquad \text{Q.E.D.}$$

This theorem shows that for the case when the expert provided information is for singleton causes, the RNOR reduces to the Noisy OR. As such, the RNOR is a mathematical generalization of Noisy OR. Because of this, we expect that it also is a generalization of the semantic mechanisms upon which Noisy OR is based.

We feel that the theorem to be proved in the next section furthers this semantic argument. This next theorem can be interpreted as showing that RNOR preserves and combines expert-provided information about how causes "work together" and does so without adding additional information. We call the ways in which causes can work together "Synergy" and "Interference".

3) Synergy and Interference: Synergy is generally agreed upon to mean a mutually advantageous conjunction of causes, that is the effect of combination of these causes is greater than the combined independent product (i.e. that presented by the Noisy OR). We use the term interference herein as the antonym of synergy, in essence when we have two or more causes that act in such a way that they interfere, the combination of these causes is less than the combined independent product. Given the potential for confusion here we paraphrase; Synergy as satisfying positive causality and the effect being greater than presented via the Noisy OR. Interference is when we satisfy positive causality and the effect is less than result of Noisy OR calculation.[3] To facilitate the ensuing discussions on Synergy and Interference, we introduce the following definition. We will later show that RNOR preserves synergy and interference.

Definition 2: Information on the probability of an effect from a combination of causes provided by an expert (or estimated via the RNOR) can be represented as a scalar multiple of the regular Noisy OR. We represent this analytically as

$$1 - p^R(\mathbf{x}) \triangleq \delta(\mathbf{x})\left(1 - p^N(\mathbf{x})\right)$$

where $p^R(\mathbf{x})$ represents the probability from a RNOR estimation and $p^N(\mathbf{x})$ represents the standard Noisy OR estimation. The factor $\delta(\mathbf{x})$ represents a scalar gain or attenuation coefficient between the two estimates.

If $\delta(\mathbf{x})$ is less than one then it represents a biased amplifying coefficient of the probability and hence synergy. Conversely if $\delta(\mathbf{x})$ is greater than one then it represents a biased attenuating coefficient and hence interference. Finally, if $\delta(\mathbf{x})$ is equal to one then an independent product combination could hold implying causal independence. The reader is referred to Appendix B for further explanation of this relationship.

Definition 2 has the following property, encapsulated in Theorem 4.

Theorem 4: When the expert/modeler does not provide $p^E(\mathbf{x})$, then for $|\mathbf{x}| \geq 2$

$$\delta(\mathbf{x}) = \prod_{i=0}^{n-1} \frac{\delta\left(\mathbf{x} \setminus \{x_i\}\right)}{\delta\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)}$$

Proof: The theorem is defined whenever $|\mathbf{x}| \geq 2$ because both $\delta(\emptyset)$ and $\delta(a)$ are defined. Since $p^E(\emptyset) = p^R(\emptyset) =$

[3]Interference $\neq$ inhibition. Inhibition occurs when positive causality is not satisfied.

$p^N(\emptyset) = 0$ we have $\delta(\emptyset) = 1$. Lemma 1 shows that, when $|\mathbf{x}| = 1$, we also have $\delta(\mathbf{x}) = 1$.

Thus, by Definition 2

$$\delta(\mathbf{x}) = \left(1 - p^R(\mathbf{x})\right) \times \frac{1}{(1 - p^N(\mathbf{x}))}.$$

By the definition of RNOR

$$\delta(\mathbf{x}) = \left[\prod_{i=0}^{n-1} \frac{\left(1 - p^R(\mathbf{x} \setminus \{x_i\})\right)}{\left(1 - p^R\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)\right)}\right] \times \frac{1}{(1 - p^N(\mathbf{x}))}.$$

Again, by Definition 2, we have the equation at the bottom of page.

By Lemma 2

$$\begin{aligned} &= \left[\prod_{i=0}^{n-1} \frac{\delta\left(\mathbf{x} \setminus \{x_i\}\right)}{\delta\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)}\right] \left(1 - p^N(\mathbf{x})\right) \\ &\quad \times \frac{1}{(1 - p^N(\mathbf{x}))} \\ &= \prod_{i=0}^{n-1} \frac{\delta\left(\mathbf{x} \setminus \{x_i\}\right)}{\delta\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)}. \end{aligned}$$

Q.E.D.

Theorems three and four suggest the value and usefulness of the RNOR. Theorem three shows that: 1) RNOR reduces to the Noisy OR when the only individual causation (singleton) probabilities are provided and 2) that RNOR allows explicit inclusion of higher order causal dependencies that an expert deems important. RNOR also combines multiple expert-provided higher order causal dependencies into still higher order probability estimates. These properties have been demonstrated in example 2.

Theorem 4 suggests how multiple expert-provided higher order causal dependencies are combined. It suggests that synergy estimates, the δ's are combined more or less as though they are independent. To see this, consider that

$$\begin{aligned} \delta(a,b,c) &= \frac{\delta(a,b)\delta(a,c)\delta(b,c)}{\delta(c)\delta(b)\delta(a)} \\ &= \frac{\delta(a,b)\delta(a,c)\delta(b,c)}{1}. \end{aligned}$$

Here we see the synergies, the δ's, combining as though they are "independent." If we look at

$$\delta(a,b,c,d) = \frac{\delta(a,b,c)\delta(a,b,d)\delta(a,c,d)\delta(b,c,d)}{\delta(b,c)\delta(a,b)\delta(a,d)\delta(c,d)}$$

we can interpret this as independent combination with "double counting" of synergies, eliminated by the terms in the denominator.

B. Computation of a CPT

Suppose that $\mathbf{x}$ contains all the parents of an event in a Bayesian Network and there are no other (unmodeled) causes. Then the computation of $p^R(\mathbf{x})$, in the course of its recursion, also computes all entries necessary for the child's conditional probability table (see example 2). Therefore, computing $p^R(\mathbf{x})$ and setting $p(y|\mathbf{x}) = p^R(\mathbf{x})$ is sufficient to specify a CPT.

If we wish, however, to consider (independent) causes not included in the model, we need to introduce the idea of a leak probability. Henrion [10] introduced the "leak probability," denoted herein as p_{leak}, which in essence represents the probability that y will be declared true even when all the listed causes are not active. Alternatively, one could surmise that the leak probability represents the causes that were considered not of interest to be explicitly modeled as causes in their own right. The idea of the leak probability is very useful as it allows the modeler to focus on the "important" and significant causes and "lump" the insignificant, incidental or unimportant causes into one factor p_{leak}. Our method of incorporating the leak probability assumes that the leak is causally independent from the causes of interest. Thus the "$y =$ true" portion of the CPT is computed using

$$(1 - p(y|\mathbf{x})) = (1 - p_{leak})\left(1 - p^R(\mathbf{x})\right). \tag{6}$$

The remainder of the CPT is determined by using the identity $p(\neg y|\mathbf{x}) = 1 - p(y|\mathbf{x})$.

Example 3: As per example 2, however, the expert also provides their estimate of leak probability to be $p_{leak} = 0.1$. Using the results determined as per Table I, (6) and the aforementioned probability identity we determine the CPT to be as per Table II.

As per any recursive technique, computational efficacy can be obtained by computing all the RNOR entries in $|\mathbf{x}|$ ascending cardinality order for the various causal combinations that correspond to $|\mathbf{x}|$ and storing the computed values for insertion rather than traversing the recursion every time. For example, when y has $n = |\mathbf{x}| = 4$ causes $\mathbf{x} = \{x_0, x_1, x_2, x_4\}$, we would calculate our estimate via the RNOR in the following order: $|\mathbf{x}| = 2$: $\{p^R(x_0,x_1), p^R(x_0,x_2), p^R(x_0,x_3), p^R(x_1,x_2), p^R(x_1,x_3)\}$ and $p^R(x_2,x_3)\}$, then $|\mathbf{x}| = 3$: $\{p^R(x_0,x_1,x_2), p^R(x_0,x_1,x_3), p^R(x_0,x_2,x_3)$ and $p^R(x_1,x_2,x_3)\}$, then $|\mathbf{x}| = 4$: $p^R(x_0,x_1,x_2,x_3)$.

Note that the order of the estimates in the curly braces is not important, just the cardinality order. The remaining step for CPT computation is to use (6).

$$\begin{aligned} &= \left[\prod_{i=0}^{n-1} \frac{\delta\left(\mathbf{x} \setminus \{x_i\}\right)\left(1 - p^N\left(\mathbf{x} \setminus \{x_i\}\right)\right)}{\delta\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)\left(1 - p^N\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)\right)}\right] \times \frac{1}{(1 - p^N(\mathbf{x}))} \\ &= \left[\prod_{i=0}^{n-1} \frac{\delta\left(\mathbf{x} \setminus \{x_i\}\right)}{\delta\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)} \prod_{i=0}^{n-1} \frac{\left(1 - p^N\left(\mathbf{x} \setminus \{x_i\}\right)\right)}{\left(1 - p^N\left(\mathbf{x} \setminus \left\{x_i, x_{(i+1)\bmod n}\right\}\right)\right)}\right] \times \frac{1}{(1 - p^N(\mathbf{x}))}. \end{aligned}$$

TABLE II
COMPUTED CPT

$\mathbf{x}$	$\emptyset$	x_0	x_1	x_2	x_3	x_0, x_1	x_0, x_2	x_0, x_3
$p(y=1\|\mathbf{x})$	0.1	0.28	0.325	0.37	0.415	0.46	0.496	0.613
$p(y=0\|\mathbf{x})$	0.9	0.72	0.675	0.63	0.585	0.54	0.504	0.387
$\mathbf{x}$	x_1, x_2	x_1, x_3	x_2, x_3	x_0, x_1, x_2	x_0, x_1, x_3	x_0, x_2, x_3	x_1, x_2, x_3	x_0, x_1, x_2, x_3
$p(y=1\|\mathbf{x})$	0.5275	0.562	0.712	0.775	0.71	0.757	0.784	0.891
$p(y=0\|\mathbf{x})$	0.4725	0.438	0.288	0.225	0.29	0.243	0.216	0.109

C. Related Work

In the proceeding section, we demonstrated how our causal theory can be used to compute CPTs. Other researchers have also attempted to address the problem of populating CPTs with researchers such as Poole [16] choosing a representation in terms of rules that provide conditional probabilities in different contexts. Poole's approach, however, can lead to an overhead in manipulating the rules. Also, this approach applies only to computing the posterior probabilities, i.e. a prior CPT is still required. Other approaches, chosen by researchers to obtain an *a-priori* CPT, include learning from a dataset [12]–[15]. Whilst these approaches have been shown to deliver remarkable performance, for example [12], one needs to recognize that with learning the accuracy of the values determined as entries for the CPT is directly related to the number of instances in the data set, i.e. the training data set requirements grow with the number of CPT entries [13]. Yet as Diez [7] put it succinctly "Very often, there is not even an incomplete database, and the knowledge engineer must resort to subjective assessment of probability obtained from human experts, who make use of their memory and of the literature published in their field". We are interested in the situations with no prior history and, furthermore where it is likely that the 'knowledge engineer' is the end user/planner.

Such cases arise when using binary Causal Analysis for plan analysis. A popular approach that has been exploited in obtaining the prior CPT is that of the Noisy OR [4] and its variants [6], [7], [9]–[11] which provide the means to represent or generate a CPT with a linear number, greater than or equal to, n entries. The principal problem with the Noisy OR is that it does not easily accommodate concepts such as synergy or interference. This is a consequence of the independence assumption of the Noisy OR.

IV. Conclusions

This paper focused on approaches that address the intractability of knowledge acquisition of CPTs in causal/Bayesian networks. We introduced the RNOR, a canonical formulation of multicausal interactions. We have shown through the development and proof of lemmas and theorems, and with examples that the RNOR allows for the building of better models by inclusion of known causal dependencies, with subsequent use of these dependencies in calculation of higher order causal probabilities. We have shown that the RNOR can subsume or reduce to the Noisy OR when only individual causation probabilities are provided. In the course of demonstrating that the RNOR permits the inclusion of causal concepts such as synergy and interference, we have also formalized a definition for these concepts and a simple means to quantify these concepts. Furthermore, we have shown that the probability estimates generated by the RNOR incorporate and factor these synergies and interferences in a meaningful way. It is through these mechanisms that the RNOR can facilitate the modeler and is applicable in cases where the Noisy OR would fail. One such example is the case where a single cause is known to have no impact on the effect unless it is used in a specific combination with other causes. This example would logically result in this cause being removed from the Noisy OR (except in the case of manual CPT intervention), hence, invalidating the use of the Noisy OR in this case, yet the RNOR elegantly accounts for this case. The development of our causal theory is embodied in the RNOR rule, we further presented a method for the determination of the CPT from this causal theory and demonstrated this by an example.

Appendix A

We commence derivation of the lower part of the RNOR rule by re-writing the Noisy OR, (3). Equation (A.1) below is such a rewriting. This can be seen by cancelling

$$1 - \vec{p}_y^{\,\mathrm{N}}(\mathbf{x}) = \frac{\left\lfloor \prod_{i=0}^{n-1} (1 - \vec{p}_y(x_i)) \right\rfloor^{n-1}}{\left\lfloor \prod_{i=0}^{n-1} (1 - \vec{p}_y(x_i)) \right\rfloor^{n-2}} \tag{A.1}$$

the numerator into the denominator. Thus, (A.1) is simply another form of our Noisy OR. The equality of (A.1) with (3) can also be seen by expanding (A.1), canceling terms, and noting that, for all $\mathbf{x}$, the factor $(1 - \vec{p}_y(x_i))$ occurs exactly once.

Expanding the right-hand side of (A.1) yields

$$1 - \vec{p}_y(\mathbf{x}) = \frac{(1 - \vec{p}_y(x_0))^{n-1} (1 - \vec{p}_y(x_1))^{n-1} \ldots}{(1 - \vec{p}_y(x_0))^{n-2} (1 - \vec{p}_y(x_1))^{n-2} \ldots} \times \frac{\ldots (1 - \vec{p}_y(x_{n-2}))^{n-1} (1 - \vec{p}_y(x_{n-1}))^{n-1}}{\ldots (1 - \vec{p}_y(x_{n-2}))^{n-2} (1 - \vec{p}_y(x_{n-1}))^{n-2}}.$$

$$
\begin{aligned}
1-\vec{p}_y(\mathbf{x}) = &\frac{(1-\vec{p}_y(x_1))\dots(1-\vec{p}_y(x_{n-2}))\,(1-\vec{p}_y(x_{n-1}))}{(1-\vec{p}_y(x_2))\,(1-\vec{p}_y(x_3))\dots(1-\vec{p}_y(x_{n-2}))\,(1-\vec{p}_y(x_{n-1}))}\\
&\times\frac{(1-\vec{p}_y(x_0))\,(1-\vec{p}_y(x_2))\dots(1-\vec{p}_y(x_{n-2}))\,(1-\vec{p}_y(x_{n-1}))}{(1-\vec{p}_y(x_0))\,(1-\vec{p}_y(x_3))\dots(1-\vec{p}_y(x_{n-2}))\,(1-\vec{p}_y(x_{n-1}))}\\
&\times\dots\times\frac{(1-\vec{p}_y(x_0))\,(1-\vec{p}_y(x_1))\dots(1-\vec{p}_y(x_{n-2}))}{(1-\vec{p}_y(x_1))\,(1-\vec{p}_y(x_2))\dots(1-\vec{p}_y(x_{n-3}))\,(1-\vec{p}_y(x_{n-2}))}.
\end{aligned}
$$

Expanding further and rearranging the factors gives the equation at the top of the page.

Note that $(1-\vec{p}_y(x_0))$ is missing from the numerator of the first quotient, $(1-\vec{p}_y(x_1))$ is missing from the numerator of the second quotient, etc. Also note that $(1-\vec{p}_y(x_0))(1-\vec{p}_y(x_1))$ is missing from the denominator of the first quotient, $(1-\vec{p}_y(x_1))(1-\vec{p}_y(x_2))$ is missing from the denominator of the second quotient, and so on.

Using a simpler notation, the last equation can be rewritten as

$$
\begin{aligned}
1-\vec{p}_y(\mathbf{x}) = &\frac{(1-\vec{p}_y\,(\mathbf{x}\setminus\{x_0\}))}{(1-\vec{p}_y\,(\mathbf{x}\setminus\{x_0,x_1\}))}\\
&\times\frac{(1-\vec{p}_y\,(\mathbf{x}\setminus\{x_1\}))}{(1-\vec{p}_y\,(\mathbf{x}\setminus\{x_1,x_2\}))}\\
&\times\dots\times\frac{(1-\vec{p}_y\,(\mathbf{x}\setminus\{x_{n-1}\}))}{(1-\vec{p}_y\,(\mathbf{x}\setminus\{x_{n-1},x_0\}))}.
\end{aligned}
$$

where $\setminus$ is used to represent the set minus. This can be compactly rewritten as

$$
\begin{aligned}
1-\vec{p}_y(\mathbf{x}) &= \frac{\prod_{i=0}^{n-1}(1-\vec{p}_y\,(\mathbf{x}\setminus\{x_i\}))}{\prod_{i=0}^{n-1}\left(1-\vec{p}_y\left(\mathbf{x}\setminus\left\{x_i,x_{(i+1)\bmod(n)}\right\}\right)\right)}\\
&=\prod_{i=0}^{n-1}\frac{(1-\vec{p}_y\,(\mathbf{x}\setminus\{x_i\}))}{\left(1-\vec{p}_y\left(\mathbf{x}\setminus\left\{x_i,x_{(i+1)\bmod(n)}\right\}\right)\right)}
\end{aligned}
$$

where $(a) \bmod (b)$ represents the modulus function.

APPENDIX B

We stated in Definition 2 that $1-p^R(\mathbf{x}) \triangleq \delta(\mathbf{x})(1-p^N(\mathbf{x}))$, where $p^R(\mathbf{x})$ represents either the information on the causal probability provided by an expert $(p^E(\mathbf{x}))$ or computed via the RNOR $(p^R(\mathbf{x}))$ for causes $\mathbf{x}$, $(1-p^N(\mathbf{x}))$ represents the standard Noisy OR as per (3). Finally, $\delta(\mathbf{x})$ represents a scalar gain or attenuation coefficient with a range $\lfloor 0, 1/(1-p^N(\mathbf{x}))\rfloor$. The only analytic exception to this is when we have $\vec{p}_y^N(\mathbf{x})=1$.

However it would seem more natural to have invoked the relationship $p^R(\mathbf{x})=\Delta(\mathbf{x})p^N(\mathbf{x})$, where $\Delta(\mathbf{x})$ represents a different scalar coefficient with range $\lfloor 0, 1/p^N(\mathbf{x})\rfloor$. The reason for saying this might be a more natural relationship is that it is directly between the provided and the calculated probability values, without any biases. Clearly if the scalar, $\Delta(\mathbf{x})>1$ then it represents an amplifying coefficient and hence synergy, conversely if $\Delta(\mathbf{x})<1$ then it represents an attenuating coefficient and hence interference. Finally, if $\Delta(\mathbf{x})=1$ we have the case where the independent product combination may hold.

It is easily shown that the scalar $\Delta(\mathbf{x})$ can be determined from $\delta(\mathbf{x})$ and $p^N(\mathbf{x})$, and vice versa

$$
\begin{aligned}
\Delta(\mathbf{x}) &= \frac{1-\delta(\mathbf{x})\left(1-\vec{p}^N(\mathbf{x})\right)}{\vec{p}^N(\mathbf{x})}, \text{or}\\
\delta(\mathbf{x}) &= \frac{1-\Delta(\mathbf{x})\vec{p}^N(\mathbf{x})}{1-\vec{p}^N(\mathbf{x})}.
\end{aligned}
\tag{B.1}
$$

Using the relationship of (B.1), when $\Delta(\mathbf{x})>1$ then $\delta(\mathbf{x})<1$, hence representing synergy, and when $\Delta(\mathbf{x})<1$ then $\delta(\mathbf{x})>1$ representing interference, and finally $\Delta(\mathbf{x})=\delta(\mathbf{x})=1$ indicates the possibility of independence.

REFERENCES

[1] P. Spirtes, C. Glymour, and R. Schienes, *Causation, Prediction and Search*. New York: Springer-Verlag, 1993, Lecture Notes in Statistics No 81.

[2] J. Pearl, *Causality: Models, Reasoning and Inference*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[3] J. F. Lemmer, "The causal Markov condition, fact or artifact?," *SIGART*, vol. 7, no. 3, p. 1996.

[4] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[5] F. V. Jensen, *An Introduction to Bayesian Networks*. New York: UCL Press and Springer-Verlag, 1996, pp. 48–52.

[6] D. Heckerman and J. S. Breese, "A new look at causal independence," in *Proc. 10th Conf. Uncertainty Artificial Intelligence*, 1994, pp. 286–292.

[7] F. J. Diez, "Parameter adjustment in Bayes networks. The generalized noisy OR gate," in *Proc. 9th Annu. Conf. Uncertainty Artificial Intelligence*, San Mateo, CA, 1993, pp. 99–105.

[8] J. Pearl, "Fusion, propagation and structuring in belief networks," *Artific. Intell.*, vol. 29, pp. 241–288, 1986.

[9] S. Srinivas, "A generalization of the noisy OR model," in *Proc. 9th Annu. Conf. Uncertainty Artificial Intelligence*, San Mateo, CA, 1993, pp. 208–215.

[10] M. Henrion, "Some practical issues in constructing belief networks," in *Uncertainty in Artificial Intelligence Three*, L. N. Kanal, T. S. Levitt, and J. F. Lemmer, Eds. Amsterdam, The Netherlands: Elsevier , 1989, pp. 161–173.

[11] P. J. F. Lucas, "Certainty-factor-like structures in Bayesian belief networks," in *Knowledge-Based Systems 14*. Amsterdam, The Netherlands: Elsevier, 2001, pp. 327–335.

[12] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A Bayesian approach to filtering junk e-mail," in *Proc. AAAI Workshop Learning Text Categorization*, Madison, WI, 1998.

[13] N. Friedman and M. Goldszmidt, "Learning Bayesian networks with local structure," in *Proc. 12th Annu. Conf. Uncertainty in Artificial Intelligence*, Portland, OR, 1996, pp. 252–262.

[14] X. Boyen and D. Koller, "Approximate learning of dynamic models," in *Proc. 12th Annu. Conf. Neural Information Processing Systems*, Denver, CO, 1998, pp. 396–402.

[15] P. Myllymaki and H. Tirri, "Constructing computationally efficient Bayesian models via unsupervised clustering," in *Probabilistic Reasoning and Bayesian Belief Networks*, A. Gammerman, Ed. Suffolk, U.K.: Alfred Waller, 1995, pp. 237–248.

[16] D. Poole, "Probabilistic partial evaluation: exploiting rule structure in probabilistic inference," in *Proc. 15th Int. Joint Conf. AI (IJCAI-97)*, Nagoya, Japan, 1997, pp. 1284–1291.

[17] I. J. Good, "A causal calculus," *British J. Philos. Sci.*, vol. 11, pp. 305–318, 1961.

[18] "A causal calculus," *British J. Philos. Sci.*, vol. 12, pp. 43–51, 1961.

John F. Lemmer received the Bachelor's degree in physics from Carnegie-Mellon University, Pittsburgh, PA, in 1960, the Master's degree in electrical engineering from Purdue University, West Lafayette, IN, in 1966, and the Ph.D. degree in computer science from the University of Maryland, College Park, in 1976.

He served in the Army from 1960 to 1970 and spent two years in Southeast Asia. He worked for various U.S. defense contractors from 1976 until 1993, when he became a civilian employee of the U.S. Air Force, Rome, NY. He has co-edited six books on uncertainty in artificial intelligence and is the co-founder of the Association for Uncertainty in Artificial Intelligence. His current research centers on the computational and human feasibility of directly applying Bayesian reasoning to medium-scale temporal causal models without overuse of independence assumptions, and without the intervention of knowledge engineers.

Don E. Gossink received the Bachelor's degree in electronic engineering in 1994 and the Ph.D. degree in information and communications technology in 1998, both from the University of South Australia, Adelaide.

He has worked for the Institute of Telecommunications Research from 1993 until 1995, when he was employed by the Defence Science and Technology Organization, Edinburgh, Australia. He has authored and co-authored over 30 refereed journal and conference papers. His current research interests are dynamic Bayesian networks, causality, artificial intelligence, and design of command and control socio-technical systems.