

Qualitative Approaches to Quantifying Probabilistic Networks

Kwalitatieve benaderingen
tot het
kwantificeren van probabilistische netwerken

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof. Dr. H.O. Voorma,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen
op maandag 12 maart 2001 des middags te 14:30 uur

door

Silja Renooij

geboren op 13 oktober 1972, te Amersfoort

promotores: Prof. Dr. J.-J. Ch. Meyer
Prof. Dr. Ir. L.C. van der Gaag
co-promotor: Dr. C.L.M. Witteman

Faculteit Wiskunde en Informatica, Universiteit Utrecht



SIKS Dissertation Series No. 2001 – 1

The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Graduate School for Information and Knowledge Systems.

ISBN 90-393-2644-4

Contents

1	Introduction	1
I	Qualitative Probabilistic Networks	9
2	Preliminaries	11
2.1	Probability theory	11
2.2	Graph theory	14
2.3	Graphical models of probabilistic independence	16
2.4	Probabilistic networks	17
3	Qualitative Probabilistic Networks	19
3.1	Defining a qualitative probabilistic network	19
3.2	Inference in a qualitative probabilistic network	37
3.3	A note on non-binary nodes	40
3.4	Discussion	43
4	Refining Qualitative Networks	45
4.1	Exploiting non-monotonic influences	46
4.2	Enhanced qualitative probabilistic networks	56
4.3	Context-specific sign-propagation	73
4.4	Pivotal pruning of trade-offs	81
4.5	Propagating multiple simultaneous observations	92
4.6	Related work	103

II	Probability Elicitation	107
5	The Elicitation Process	109
5.1	The elicitation process	110
5.2	Presentation	115
5.3	Methods	117
5.4	Discussion	123
6	Designing a New Elicitation Method	125
6.1	Modes of probability expression: previous studies	126
6.2	Design considerations and goals	128
6.3	Our study	129
6.4	The new elicitation method	142
6.5	Conclusions	144
7	Experiences with our Elicitation Method	145
7.1	Initial experiences with probability elicitation	146
7.2	Evaluation of the elicitation method	147
7.3	Evaluation of the elicited probabilities	150
7.4	Concluding observations	157
8	Conclusions	159
8.1	A qualitative approach to probabilistic reasoning	159
8.2	A qualitative approach to probability elicitation	164
8.3	Directions for future research	165
A	The Oesophagus Network	171
B	Statistical Methods	177
C	Questionnaires and Evaluation Forms	183
	Bibliography	191
	Samenvatting	203
	Dankwoord	207
	Curriculum Vitae	209
	List of Notations	211

CHAPTER 1

Introduction

Probabilistic networks have become widely accepted as practical representations of knowledge for reasoning under uncertainty and, more specifically, for decision support. The framework of probabilistic networks combines a graphical representation of a domain's variables and the relations between them, with probabilities that represent the uncertainties in the domain [90]. The framework offers powerful algorithms for reasoning with these probabilities in a mathematically correct way. These algorithms allow for causal reasoning, from cause to effect, for diagnostic reasoning, from effect to cause, and for case-specific reasoning. In a medical context, for example, case-specific reasoning amounts to taking data of a specific patient into account when computing the probability of some outcome.

Applications of probabilistic networks can be found in areas such as (medical) diagnosis and prognosis, planning, monitoring, vision, information retrieval, natural language processing, and e-commerce. Examples of fielded systems (see [49]) include Intel's processor fault diagnosis system; General Electric's generator monitoring system; a real-time weapon scheduling system for the US Navy; the e-commerce product Frontmind; the Vista system used at NASA Mission Control Center to interpret live telemetry and provide advice on the likelihood of failures of the space shuttle's propulsion systems; and the TRACS system used by the UK's Defence Research and Evaluation Agency for predicting the reliability of military vehicles. The most widely used probabilistic networks, however, are probably the ones embedded in Microsoft products, including the Answer Wizard of Office 95, the Office Assistant (paperclip) of Office 97, and a number of technical support troubleshooters.

The list of fielded applications, although impressive, reveals that the majority of probabilistic network applications are developed for, and no doubt funded by, major industries, and mostly concern technical domains, describing the (dis)functioning of machines. In such domains the relationships between the variables are mostly deterministic and, as a result, there are very few uncertainties. Traditionally, probabilistic networks, and knowledge-based systems in general,

are more suited to nature-inspired domains such as meteorology, agriculture, and most notably, medical diagnosis. Examples of scientific applications include the Hailfinder weather forecasting system; a network for mildew management in wheat; the BOBLO network for blood typing cattle; the MUNIN system for interpreting electromyocardiograms; and Pathfinder for lymph-node pathology diagnosis. These and other examples of scientific applications that build on the framework of probabilistic networks are described in e.g. [1, 2, 4, 30, 54, 61, 62, 68, 70, 78, 91, 95, 113]. For most of these applications only prototype systems have been demonstrated and very few are known to have been implemented in practice. The main reason for this observation is probably that building a probabilistic network is a difficult and time-consuming task. The networks are typically constructed with the help of experts in the domain of application and the more complex and uncertain the domain, the harder the task of building the network and the more time the experts involved are required to invest in the project.

Building a probabilistic network for an application domain basically involves three tasks. The first task is to identify the important domain variables and their possible values. The second task is to identify the relationships between these variables. The variables and their relationships are expressed in a directed acyclic graph, with nodes modelling variables and arcs modelling relationships between the variables. The resulting graph is referred to as the network's qualitative part. As the final task, the probabilities that constitute the network's quantitative part are to be obtained; local (conditional) probability distributions are required for each variable in the network. In principle, the three tasks are performed sequentially. However, as with any large system, the design and construction of a probabilistic network often follows a spiral life-cycle model, iterating over the tasks until a satisfactory network results [79].

Well-known knowledge-acquisition and engineering techniques for identifying domain variables, their values, and the relationships between them, can be employed, to at least some extent, for the construction of the qualitative part of a probabilistic network. Identifying the important domain variables and their values is typically performed with the help of one or more experts. The variables are to be expressed as statistical variables with a set of mutually exclusive and collectively exhaustive values. The meanings of the resulting statistical variables and their values have to be properly documented to avoid any ambiguity in future reference. The relationships between the variables can be either elicited from the domain experts or learned from data. In eliciting the relationships from experts, the concept of causality can be used as a heuristic guiding principle; in graphical terms, the direction of causality is used for directing the arcs between related variables [55]. In data-rich applications, data collections that are large, up-to-date and reliable can be used to automatically learn the graphical structure of a probabilistic network [14].

Although the construction of the qualitative part of a probabilistic network requires considerable effort, it is generally considered feasible. The qualitative part provides a graphical representation of the problem domain that allows for easy communication with experts; even experts who know little about probability theory or the framework of probabilistic networks are able to interpret and refine the qualitative part of a network. Quantification, however, is considered a far harder task and is, in fact, often referred to as a major obstacle in building a probabilistic network [37, 61]. Quantifying probabilistic networks is the focus of this thesis; more in particular, we will discuss *qualitative* approaches to quantifying probabilistic networks.

Real-life probabilistic networks may comprise tens or hundreds of variables and can easily require thousands of probabilities. In most problem domains, various sources of probabilistic information are available that seem usable for quantification of a network. Examples of such sources are (statistical) data, literature, and human experts. Unfortunately, these sources seldom provide for ready-made probability assessments. To allow for distilling probability assessments from data collections, for example, these collections should be up to date and unbiased. Also, the variables and values that are recorded in the data should match those modelled in the network. In addition, the collection should be large enough to allow for reliable assessments for the required probabilities. As for each variable several probability distributions, conditional on the values of the variable's parents in the network's graphical part, need to be specified, only small subsets of the data can be used to estimate the specific probabilities required. In insufficiently large data collections, these subsets may be empty or too small to allow for meaningful assessments. To conclude, a data collection should not have too many missing values. Missing values are due either to errors or to omission. A value for the result of a diagnostic test that is not performed, for example, will not be recorded in the data. If the missing values distributed over the data in a non-random way, they can easily introduce biases in the estimated probabilities.

Literature often provides abundant probabilistic information. For example, the sensitivity and specificity of medical diagnostic tests, as well as their typical ranges, are often reported in medical handbooks or journals. However, as probabilistic information found in literature is often derived from a population with specific characteristics, care should be taken not to use this information for populations with other characteristics. Another problem with probabilistic information reported in literature is that it is seldom directly amenable to encoding in a probabilistic network: conditional probabilities are sometimes given in a direction opposite to the direction required for the network, and the information is often incomplete. For example, medical literature often reports conditional probabilities for the presence of symptoms given a disease, but not for these symptoms occurring in the absence of the disease. In addition, probabilistic networks often contain hidden variables, that is, variables for which no value can be observed in the physical world; probabilistic information concerning such variables will be lacking altogether.

Another source of probabilistic information is the knowledge and experience of domain experts. The importance of this source in the construction of the quantitative part of a probabilistic network should not be underestimated. An expert's knowledge and experience can help, not just in assessing the probabilities required, but also in fine-tuning probabilities obtained from other sources to the specifics of the domain at hand, and in verifying them within the context of the network. Unfortunately, experts are often uncomfortable with having to provide probabilities. Moreover, the problems of bias encountered when directly eliciting probabilities from experts are widely known [64]. An expert's assessments, for example, may not be properly calibrated and may reflect various biases resulting from the heuristics, or efficient shortcuts, that experts, often unconsciously, use for the assessment task. Examples of such biases are overestimation, where an expert consistently gives probability assessments that are higher than the true probabilities, and overconfidence, where assessments for likely events are too high and assessments for unlikely events are too low. In the field of decision analysis various methods have been developed to counteract biases during probability elicitation [83, 129]. These methods, however, are often complicated and their use is very time-consuming; the methods are therefore unsuitable

for eliciting a large number of probabilities from experts whose time is a scarce and expensive commodity.

Expert judgement is considered the least objective and least accurate source of probabilistic information. However, as other sources of probabilistic information seldom provide for all required probabilities, the knowledge and experience of a domain expert is the single remaining source that can be exploited for quantifying a probabilistic network [61]. The probabilities with which a network is quantified are therefore necessarily *subjective*, describing the state of an expert's knowledge and beliefs [110]. By assuming that probabilities are subjective by nature, it is in principle possible for a domain expert to give an assessment for the likelihood of any event, even if he¹ knows little about it.

Due to the incompleteness of probabilistic information from data and from literature, and as a result of partial domain knowledge and biases, the numbers obtained for a probabilistic network are inevitably inaccurate. Whether or not these inaccuracies are problematic, depends on the extent to which they influence the behaviour of the network. In general, the robustness of the graphical structure of a probabilistic network, reflecting the independence and relevance relationships between its variables, is considered more crucial than the individual probabilities [35, 124]. The accuracy of the individual probabilities will, nonetheless, influence the network's output. To investigate the possible effects of the inaccuracies in the network's probabilities, sensitivity and uncertainty analyses can be performed [83]. In an uncertainty analysis, the assessments of all conditional probabilities are varied simultaneously by drawing, for each of them, a value from a pre-specified distribution; an uncertainty analysis reveals the overall reliability of the network's output. Uncertainty analysis of a large real-life probabilistic network for liver and biliary disease has led to the suggestion that probabilistic networks are highly insensitive to inaccuracies in their probability assessments [57, 92]. However, in this particular analysis, the average effect of varying the conditional probabilities was studied, whereas the variance in the effect, rather than their average value, truly reflects the effects of inaccuracies [25]. In a sensitivity analysis, the assessments of one or more conditional probabilities are varied simultaneously over a plausible interval; a sensitivity analysis yields insight into the separate effects of conditional probabilities on the network's output. Sensitivity analysis of a real-life probabilistic network for congenital heart disease revealed large effects of varying conditional probabilities [24]. To be able to draw any decisive conclusions about the effects of inaccuracies, more research is required; it seems likely, though, that these effects will vary from application to application.

As domain experts are often the only source of probabilistic information, and probability elicitation from experts is known to be problematic, the probabilities obtained from experts should be taken as rough initial assessments. These rough assessments can be used as a starting point in an iterative procedure aimed at refining the assessments where necessary. As sensitivity analysis reveals the effect of varying a single conditional probability on the network's output, it can be used in such an iterative procedure [26]. In the first step, a network is quantified with initial assessments. Then a sensitivity analysis of the network is performed, upon which the most influential probabilities are refined. Iteratively performing sensitivity analyses and refining probabilities is pursued until the network's behaviour is satisfactory, or until higher accuracy can no longer be attained due to lack of resources.

¹ Anywhere we use a masculine pronoun, the feminine form is obviously understood to be included.

Aim of the thesis

In this thesis, we address the quantification of probabilistic networks with the help of domain experts. As probabilistic networks become more popular, they are being applied to problems of increasing size and complexity. The design of methods tailored to fast and easy elicitation of large numbers of probabilities is therefore becoming increasingly important. For the first steps in a step-wise refinement procedure, we feel that the experts should be accommodated by allowing them to express uncertainties in any format they feel comfortable with [36]. These formats may be quantitative in nature, such as point estimates or probability intervals, but may also be more qualitative, such as verbal expressions of uncertainty or statements regarding the influence of one variable on another variable. In this thesis, we propose two different qualitative approaches that can be exploited in quantifying probabilistic networks.

In the first approach, we propose to quantify probabilistic networks with purely qualitative statements, resulting in *qualitative probabilistic networks*. These qualitative networks allow for probabilistic reasoning in a qualitative way. For some application domains, probabilistic reasoning with a qualitative network may be specific enough. In domains that require a more fine-grained level of detail, qualitative probabilistic reasoning can be used for studying the projected network's reasoning behaviour, prior to the assessment of the required probabilities. Then, if the network's graphical part is considered robust, the qualitative statements can be used as constraints on the conditional probability distributions to be assessed [36, 77]. For fast elicitation of initial assessments for the required probabilities from experts, we propose adding qualitative ingredients to a method for direct elicitation of numbers.

A qualitative probabilistic network is a qualitative abstraction of a probabilistic network with the same graphical part [134]. In a quantified probabilistic network, the conditional probability distributions specified in essence capture the directions and strengths of the influences between related variables. In a qualitative probabilistic network, in contrast, only the directions of the influences are captured. These directions are indicated by qualitative signs, which are elicited from domain experts. To this end, the experts are required to give statements of stochastic dominance, that is, statements such as “the larger the tumour, the more likely it is that there are metastases”; expressing their knowledge in such statements requires considerably less effort from the experts than the specification of numbers [33]. As the stochastic dominance statements have a mathematical foundation, it is possible to reason about these statements in a mathematically correct way, thereby providing for probabilistic reasoning in a purely qualitative way. As stochastic dominance statements concern entire probability distributions, they can not be directly translated into separate probabilities. They can, however, be used as constraints on the distributions to be assessed.

In a qualitative probabilistic network, the probabilistic relations between the variables are modelled at a very coarse level of detail. As only the direction of influence between two variables is modelled and there is no notion of strength, reasoning with a qualitative probabilistic network often leads to uninformative results. As we envision an important role for qualitative probabilistic networks in the construction of quantitative probabilistic networks, it is necessary to derive as much information from them as possible. To this end, we will discuss a number of refinements of the basic formalism of qualitative probabilistic networks. These refinements, such as adding a notion of strength and of context, lead to more informative results during qualitative reasoning

and thus allow for more effectively studying the projected network's reasoning behaviour. In addition, the refinements provide stronger constraints on the network's quantification.

For initial quantification of a probabilistic network, the rate at which probabilities are assessed is more important than their accuracy. For actually obtaining probabilities from domain experts, we therefore propose using a second qualitative approach. We designed an elicitation method that is easy to use and understand, thereby allowing for fast assessment of large numbers of probabilities. In the new method we use a probability scale, which is a well-established, easy to understand elicitation aid. As a probability scale requires from experts the uncomfortable task of having to state numbers, we augmented the scale with verbal probability expressions such as *certain* and *probable*, thereby allowing experts to express uncertainties in either verbal or numerical form. Other ingredients of our method are the use of text-fragments for representing the required probabilities, and the grouping of probabilities that have to sum to one.

To summarise, we propose two qualitative approaches to quantifying probabilistic networks. First, we propose to quantify a network with qualitative signs; this results in a qualitative probabilistic network for studying the projected network's reasoning behaviour and in a set of constraints on the required probabilities. Subsequently, we propose to use our elicitation method to quantify the network with actual probabilities elicited from experts. From these two qualitative approaches, an initial rough quantification can be obtained, which can then serve as a starting point for the iterative refinement procedure that we outlined above.

Acknowledging the importance of developing methods to aid the construction of probabilistic networks in general, and their quantification more specifically, the objectives of this thesis are, in short

- to refine the basic formalism of qualitative probabilistic networks in order to arrive at more informative results during qualitative reasoning and, hence, at more insight into constraints on the network's quantification;
- to design an elicitation method, tailored to the fast and easy elicitation of probabilities, which allows the use of both verbal and numerical probability expressions.

Outline of the thesis

The thesis is divided into two parts corresponding to the two objectives mentioned above. In the first part we describe the results of our studies into refining the framework of qualitative probabilistic networks. To this end, we first present some preliminaries concerning graph theory, probability theory and probabilistic networks in Chapter 2. The basic framework of qualitative probabilistic networks is reviewed in Chapter 3. This framework was first introduced by Wellman [134] and later extended by Henrion and Druzdzel [33, 34, 56]; our review will be more detailed and more formal than previous presentations so as to provide a solid basis for the refinements of the framework of qualitative probabilistic networks we will propose in Chapter 4. In Chapter 4, we will discuss refining the level of detail of the set of qualitative properties of a qualitative probabilistic network by adding, for example, notions of strength and context. In addition,

we will discuss extending the standard algorithm for reasoning with a qualitative network in order to arrive at more informative results. Also, we will propose a new algorithm for isolating troublesome parts of the network. Parts of this chapter also appeared in [100], [102], [103], [104], and [105].

In the second part of the thesis, we study the design and evaluation of an elicitation method that combines both verbal and numerical expressions of uncertainty. To this end, Chapter 5 provides a general discussion of issues concerning probability elicitation. This work is also presented in [98]. The discussion from Chapter 5 serves as a starting point for the design of our new elicitation method, that is tailored to the fast and easy elicitation of a large number of probabilities. The experimental studies into the use of verbal expressions of probability underlying our method are discussed in Chapter 6, which is a revised version of [106]; this chapter is concluded with a description of the method. The new elicitation method was used with two oncologists from the Netherlands Cancer Institute, Antoni van Leeuwenhoekhuis, to quantify a real-life probabilistic network for oesophageal carcinoma. Our experiences, and those of the experts, with the use of the method are described in Chapter 7; this chapter also discusses an evaluation study of the performance of the resulting network. The work of this chapter appeared partially in [125] and [126].

The thesis is concluded with Chapter 8, which provides a summary of presented results and some directions for further research. All illustrative examples used throughout the thesis are highly simplified fragments of the probabilistic network for oesophageal carcinoma; Appendix A presents a detailed explanation of the network.

Part I

Qualitative Probabilistic Networks

In which we study enhancements of the framework of qualitative probabilistic networks. Qualitative probabilistic networks allow for studying the reasoning behaviour of an, as yet, unquantified probabilistic network. Once the network exhibits satisfactory qualitative behaviour, it reveals constraints on the probability distributions that are required for its quantitative part. As qualitative probabilistic networks can play an important role in the construction of a probabilistic network and in its quantification more specifically, it is important to make the formalism as expressive as possible.

Preliminaries

In this chapter, we review the basic concepts from probability theory and graph theory that are required to understand the ensuing chapters. In addition, probabilistic networks, as graphical representations of a probability distribution, are reviewed.

2.1 Probability theory

In this section, we will briefly review the basic concepts from probability theory; more details can be found in, for example, [107].

We consider a set of variables U . We assume throughout this thesis that all variables can take on one value from an exhaustive set of possible values; we refer to these variables as statistical variables. We will mostly denote variables by uppercase letters, where letters from the end of the alphabet denote sets of variables and letters from the beginning of the alphabet denote a single variable; lowercase letters are used to indicate a value, or combination of values, for a variable or set of variables, respectively. When the value of a variable is known, we say that the variable is *instantiated* or *observed*; we sometimes refer to the observed value as *evidence*.

Let $A \in U$ be a statistical variable that can have one of the values a_1, \dots, a_m , $m \geq 1$. The *value statement* $A = a_i$, stating that variable A has value a_i , can be looked upon as a logical proposition having either the truth-value true or false. A combination of value statements for a set of variables then is a logical conjunction of atomic propositions for the separate variables from that set. A proposition $A = a_i$ is also written a_i for short; a combination of value statements for the separate variables from a set X is written x for short. The set of variables U can now be looked upon as spanning a *Boolean algebra* of such logical propositions. We recall that a Boolean algebra \mathcal{B} is a set of propositions with two binary operations \wedge (conjunction) and \vee (disjunction), a unary operator \neg (negation), and two constants \mathbf{F} (false) and \mathbf{T} (true) which

behave according to logical truth tables. For any two propositions a and b we will often write ab instead of $a \wedge b$; we will further write \bar{a} for $\neg a$.

A *joint probability distribution* now is defined as a function on a set of variables spanning a Boolean algebra of propositions.

Definition 2.1 (joint probability distribution) *Let \mathcal{B} be the Boolean algebra of propositions spanned by a set of variables U . Let $\text{Pr} : \mathcal{B} \rightarrow [0, 1]$ be a function such that*

- for all $a \in \mathcal{B}$, we have $\text{Pr}(a) \geq 0$, and $\text{Pr}(F) = 0$, more specifically,
- $\text{Pr}(T) = 1$, and
- for all $a, b \in \mathcal{B}$, if $a \wedge b \equiv F$ then $\text{Pr}(a \vee b) = \text{Pr}(a) + \text{Pr}(b)$.

Then, Pr is called a joint probability distribution on U .

A joint probability distribution Pr on a set of variables U is often written as $\text{Pr}(U)$. For each proposition $a \in \mathcal{B}$, the function value $\text{Pr}(a)$ is termed the *probability* of a . We now introduce the concept of *conditional probability*.

Definition 2.2 (conditional probability) *Let Pr be a joint probability distribution on a set of variables U and let $X, Y \subseteq U$. For any combination of values x for X and y for Y , with $\text{Pr}(y) > 0$, the conditional probability of x given y , denoted as $\text{Pr}(x \mid y)$, is defined as*

$$\text{Pr}(x \mid y) = \frac{\text{Pr}(xy)}{\text{Pr}(y)}.$$

The conditional probability $\text{Pr}(x \mid y)$ expresses the amount of certainty concerning the truth of x given that y is known *with certainty*. Throughout this thesis, we will assume that all conditional probabilities $\text{Pr}(x \mid y)$ we specify, are properly defined, that is, $\text{Pr}(y) > 0$. The conditional probabilities $\text{Pr}(x \mid y)$ for all propositions x once more constitute a joint probability distribution on U , called the *conditional probability distribution given y* .

We will now state some convenient properties of probability distributions.

Proposition 2.3 (chain rule) *Let Pr be a joint probability distribution on a set of variables $U = \{A_1, \dots, A_n\}$, $n \geq 1$. Let a_i represent a value statement for variable $A_i \in U$, $i = 1, \dots, n$. Then, we have*

$$\text{Pr}(a_1 \dots a_n) = \text{Pr}(a_n \mid a_1 \dots a_{n-1}) \cdot \dots \cdot \text{Pr}(a_2 \mid a_1) \cdot \text{Pr}(a_1).$$

Another useful property is the *marginalisation property*.

Proposition 2.4 (marginalisation) *Let Pr be a joint probability distribution on a set of variables U . Let $A \in U$ be a variable with possible values a_i , $i = 1, \dots, m$; let $X \subseteq U$. Then, the probabilities*

$$\text{Pr}(x) = \sum_{i=1}^m \text{Pr}(x \wedge a_i)$$

for all combinations of values x for X define a joint probability distribution on X .

The joint probability distribution $\Pr(X)$ defined in the previous proposition is called the *marginal probability distribution* on X .

From the marginalisation property follows the *conditioning property* for conditional probability distributions.

Proposition 2.5 (conditioning) *Let \Pr be a joint probability distribution on a set of variables U . Let $A \in U$ be a variable with values a_i , $i = 1, \dots, m$; let $X \subseteq U$. Then,*

$$\Pr(x) = \sum_{i=1}^m \Pr(x | a_i) \cdot \Pr(a_i)$$

for all combinations of values x for X .

The following theorem is known as Bayes' rule and may be used to reverse the 'direction' of conditional probabilities.

Theorem 2.6 (Bayes' rule) *Let \Pr be a joint probability distribution on a set of variables U . Let $A \in U$ be a variable with $\Pr(a) > 0$ for some value a of A ; let $X \subseteq U$. Then,*

$$\Pr(x | a) = \frac{\Pr(a | x) \cdot \Pr(x)}{\Pr(a)}$$

for all combinations of values x for X with $\Pr(x) > 0$.

The following definition captures the concept of independence.

Definition 2.7 (independence) *Let \Pr be a joint probability distribution on a set of variables U and let $X, Y, Z \subseteq U$. Let x be a combination of values for X , y for Y and z for Z . Then, x and y are called (mutually) independent in \Pr if*

$$\Pr(xy) = \Pr(x) \cdot \Pr(y);$$

otherwise, x and y are called dependent in \Pr . The propositions x and y are called conditionally independent given z in \Pr if

$$\Pr(xy | z) = \Pr(x | z) \cdot \Pr(y | z);$$

otherwise, x and y are called conditionally dependent given z in \Pr .

The set of variables X is conditionally independent of the set of variables Y given the set of variables Z in \Pr , denoted as $I_{\Pr}(X, Z, Y)$, if for all combinations of values x, y, z for X, Y and Z , respectively, we have

$$\Pr(x | yz) = \Pr(x | z);$$

otherwise, X and Y are called conditionally dependent given Z in \Pr .

Note that the above propositions and definitions are explicitly stated for all value statements for a certain variable, or for all combinations of values for a set of variables. From here on, we will often use a more schematic notation that refers only to the variables involved; this notation implicitly states a certain property to hold for *any* value statement or combination of value statements for the involved variables. For example, using a schematic notation, the chain rule would be written as

$$\Pr(A_1 \dots A_n) = \Pr(A_n | A_1 \dots A_{n-1}) \cdot \dots \cdot \Pr(A_2 | A_1) \cdot \Pr(A_1).$$

If all variables have m possible values, then this schema represents m^n different equalities.

2.2 Graph theory

In this section we review some graph-theoretical notions; more details can be found in, for example, [53]. Generally, graph theory distinguishes between two types of graph: directed and undirected graphs. We will only discuss directed graphs, or *digraphs* for short.

Definition 2.8 (digraph) A directed graph G is a pair $G = (V(G), A(G))$, where $V(G)$ is a finite set of nodes and $A(G)$ is a set of ordered pairs (V_i, V_j) , $V_i, V_j \in V(G)$, called arcs.

When interested in only part of a digraph, we can consider a *subgraph*.

Definition 2.9 (subgraph) Let G be a digraph. A graph $H = (V(H), A(H))$ is a subgraph of G if $V(H) \subseteq V(G)$ and $A(H) \subseteq A(G)$. A subgraph H of G is a full subgraph if $A(H) = A(G) \cap (V(H) \times V(H))$; the full subgraph H is said to be induced by $V(H)$.

We will often write $V_i \rightarrow V_j$ for (V_i, V_j) to denote an arc from node V_i to node V_j in G .

Arcs entering into or emanating from a node are said to be *incident* on that node. The numbers of arcs entering into or emanating from a node are termed the *in-degree* and the *out-degree* of the node, respectively.

Definition 2.10 (degree) Let G be a digraph. Let V_i be a node in G . Then, the in-degree of V_i equals the number of nodes $V_j \in V(G)$ for which $V_j \rightarrow V_i \in A(G)$. The out-degree of V_i equals the number of nodes $V_j \in V(G)$ for which $V_i \rightarrow V_j \in A(G)$. The degree of V_i equals the sum of its in-degree and out-degree.

The following definition pertains to the family members of a node in a digraph.

Definition 2.11 (family) Let G be a digraph. Let V_i, V_j be nodes in G . Node V_j is a parent of node V_i if $V_j \rightarrow V_i \in A(G)$; node V_i then is a child of V_j . The set of all parents of V_i is written as $\pi_G(V_i)$; its children are denoted by $\sigma_G(V_i)$. An ancestor of node V_i is a member of the reflexive and transitive closure of the set of parents of node V_i ; the set of all ancestors of V_i is denoted $\pi_G^*(V_i)$. A descendant of node V_i is a member of the reflexive and transitive closure of the set of children of node V_i ; the set of descendants of V_i is denoted $\sigma_G^*(V_i)$.

The set $\pi_G(V_i) \cup \sigma_G(V_i) \cup \pi_G(\sigma_G(V_i))$ of node V_i is called the Markov blanket of V_i .

As long as no ambiguity can occur, the subscript G from π_G etc. will often be dropped.

Arcs in a digraph model relationships between two nodes. Relationships between more than two nodes are modelled with *hyperarcs*.

Definition 2.12 (hyperarc) Let G be a digraph. A hyperarc in G is an ordered pair (V', V_i) with $V' \subseteq V(G)$ and $V_i \in V(G)$.

From a digraph, sequences of nodes and arcs can be read.

Definition 2.13 (simple/composite trail) Let G be a digraph. Let $\{V_0, \dots, V_k\}$, $k \geq 1$, be a set of nodes in G . A trail t from V_0 to V_k in G is an alternating sequence $V_0, A_1, V_1, \dots, A_k, V_k$ of nodes and arcs $A_i \in A(G)$, $i = 1, \dots, k$, such that $A_i \equiv V_{i-1} \rightarrow V_i$ or $A_i \equiv V_i \rightarrow V_{i-1}$ for every two successive nodes V_{i-1}, V_i in the sequence; k is called the length of the trail t . A trail t is simple if V_0, \dots, V_{k-1} are distinct; otherwise the trail is termed composite.

We will often write $V_i \in V(t)$ to denote that V_i is a node on trail t ; the set of arcs on the trail will be denoted by $A(t)$.

Definition 2.14 (cycle) *Let G be a digraph. Let V_0 be a node in G and let t be a simple trail from V_0 to V_0 in G with length one or more. The trail t is a cycle if $V_{i-1} \rightarrow V_i \in A(t)$ for every two successive nodes V_{i-1}, V_i on t .*

Note that a simple trail can be a cycle, but never contains a subtrail that is a cycle.

Throughout this thesis we will assume all digraphs to be *acyclic*, unless stated otherwise.

Definition 2.15 ((a)cyclic digraph) *Let G be a digraph. G is called cyclic if it contains at least one cycle; otherwise it is called acyclic.*

We also assume that all digraphs are *connected*, unless explicitly stated otherwise.

Definition 2.16 ((un)connected digraph) *Let G be a digraph. G is connected if there exists at least one simple trail between any two nodes from $V(G)$; otherwise G is unconnected.*

Sometimes the removal of a single node along with its incident arcs will cause a connected digraph to become unconnected. Such a node is called an *articulation node* or *cut-vertex*.

Definition 2.17 (articulation node) *Let G be a connected digraph. Node $V_i \in V(G)$ is an articulation node for G if the subgraph of G induced by $V(G) \setminus \{V_i\}$ is unconnected.*

In subsequent chapters we will require an extended definition for a trail in a digraph. We will build upon the observation that a simple trail in G , as defined previously, forms a subgraph of G .

Definition 2.18 (trail) *Let G be a digraph. Then, $t = ((V(t), A(t)), V_0, V_k)$ is a trail from V_0 to V_k in G if $(V(t), A(t))$ is a connected subgraph of G and $V_0, V_k \in V(t)$.*

Note that as a trail t comprises a digraph, we can define a *subtrail* of t as comprising a connected subgraph of t . We will often consider simple trails that specify no more than one incoming arc for each node on the trail.

Definition 2.19 (sinkless trail) *Let G be a digraph. Let V_0, V_k be nodes in G and let t be a simple trail from V_0 to V_k in G . A node $V_i \in V(t)$ is called a head-to-head node on t , if $V_{i-1} \rightarrow V_i \in A(t)$ and $V_i \leftarrow V_{i+1} \in A(t)$ for the three successive nodes V_{i-1}, V_i, V_{i+1} on t . If no node in $V(t)$ is a head-to-head node, the simple trail t is called a sinkless trail. A composite trail t from V_0 to V_k is sinkless if all simple trails from V_0 to V_k in t are sinkless.*

We will now define three operations on trails for determining the inverse of a trail, the concatenation of trails, and the parallel composition of trails.

Definition 2.20 (inverse trail) *Let G be a digraph and let V_0, V_k be nodes in G . Let $t = ((V(t), A(t)), V_0, V_k)$ be a trail from V_0 to V_k in G . Then, the trail inverse t^{-1} of t is the trail $((V(t), A(t)), V_k, V_0)$ from V_k to V_0 in G .*

The trail concatenation of two trails in G is again a trail in G .

Definition 2.21 (trail concatenation) Let G be a digraph and let V_0, V_k and V_m be nodes in G . Let $t_i = ((V(t_i), A(t_i)), V_0, V_k)$ and $t_j = ((V(t_j), A(t_j)), V_k, V_m)$ be trails in G from V_0 to V_k , and from V_k to V_m , respectively. Then, the trail concatenation $t_i \circ t_j$ of trail t_i and trail t_j is the trail $((V(t_i) \cup V(t_j), A(t_i) \cup A(t_j)), V_0, V_m)$ from V_0 to V_m .

Similarly, the parallel trail composition of two trails in G is again a trail in G .

Definition 2.22 (parallel trail composition) Let G be a digraph and let V_0, V_k be nodes in G . Let $t_i = ((V(t_i), A(t_i)), V_0, V_k)$ and $t_j = ((V(t_j), A(t_j)), V_0, V_k)$ be two trails from V_0 to V_k in G . Then, the parallel trail composition $t_i \parallel t_j$ of trail t_i and trail t_j is the trail $((V(t_i) \cup V(t_j), A(t_i) \cup A(t_j)), V_0, V_k)$ from V_0 to V_k .

Note that the subgraph constructed for the trail that results from the trail concatenation or parallel trail composition of two trails is the same for both operations. However, as the arguments indicating the beginning and end of a trail are treated differently, we consider these to be two different kind of operations.

2.3 Graphical models of probabilistic independence

Graph theory and probability theory meet in the framework of *graphical models*. This framework allows for the representation of probabilistic independence by means of a graphical structure in which the nodes represent variables and lack of arcs indicates conditional independence. For the purpose of this thesis, we only consider directed graphical models; more information on graphical models can be found in [72].

The probabilistic meaning that is assigned to a digraph builds upon the concepts of *blocked trail* and *d-separation*. The definitions provided here are enhancements of the original definitions presented in [90], based upon recent insights [124].

Definition 2.23 (blocked and active trail) Let $G = (V(G), A(G))$ be an acyclic digraph and let A, B be nodes in G . A simple trail $t = ((V(t), A(t)), A, B)$ from A to B in G is blocked by a set of nodes $X \subseteq V(G)$ if (at least) one of the following conditions holds:

- $A \in X$ or $B \in X$;
- there exist nodes $C, D, E \in V(t)$ such that $D \in X$ and $D \rightarrow C, D \rightarrow E \in A(t)$;
- there exist nodes $C, D, E \in V(t)$ such that $D \in X$ and $C \rightarrow D, D \rightarrow E \in A(t)$;
- there exist nodes $C, D, E \in V(t)$ such that $C \rightarrow D, E \rightarrow D \in A(t)$ and $\sigma^*(D) \cap X = \emptyset$.

Otherwise, the trail t is called active with respect to X .

When every trail between two nodes is blocked, the nodes are said to be *d-separated* from each other.

Definition 2.24 (d-separation) Let G be an acyclic digraph and let X, Y, Z be sets of nodes in G . The set Z is said to d-separate X from Y in G , written $\langle X \mid Z \mid Y \rangle_G^d$, if every simple trail in G from a node in X to a node in Y is blocked by Z .

The framework of graphical models relates the nodes in a digraph to the variables in a probability distribution. To this end, each variable considered is represented by a node in the digraph, and vice versa. As the set of variables is equivalent to the set of nodes, throughout this thesis we will no longer make an explicit distinction between nodes and variables: if we say that a node has a value, we mean that the variable associated with that node has that value. Conditional independence is captured by the arcs in the digraph by means of the d-separation criterion: nodes that are d-separated in the digraph are associated with conditionally independent variables. The relation between graph theory and probability theory in a directed graphical model now becomes apparent from the notion of (directed) *independence map* (I-map).

Definition 2.25 (independence map) *Let G be an acyclic digraph and let \Pr be a joint probability distribution on $V(G)$. Then, G is called an independence map, or I-map for short, for \Pr if*

$$\langle X | Z | Y \rangle_G^d \Rightarrow \Pr(X | YZ) = \Pr(X | Z),$$

for all sets of nodes $X, Y, Z \subseteq V(G)$.

An I-map is a digraph having a special meaning: nodes that are not connected by an arc in the digraph correspond to variables that are independent in the represented probability distribution; nodes that are connected, however, need not necessarily represent dependent variables. For further details, the reader is referred to [90].

2.4 Probabilistic networks

Probabilistic networks are graphical models supporting the modelling of uncertainty in large complex domains. In this section we briefly review the framework of probabilistic networks, also known as (Bayesian) belief networks, Bayes nets, or causal networks [90].

The framework of probabilistic networks was designed for *reasoning with uncertainty*. The framework is firmly rooted in probability theory and offers a powerful formalism for representing a joint probability distribution on a set of variables. In the representation, the knowledge about the independences between the variables in a probability distribution is explicitly separated from the numerical quantities involved. To this end, a probabilistic network consists of two parts: a qualitative part and an associated quantitative part. The qualitative part of a probabilistic network takes the form of an acyclic directed graph G . Informally speaking, we take an arc $A \rightarrow B$ in G to represent a causal relationship between the variables associated with the nodes A and B , designating B as the effect of the cause A . Absence of an arc between two nodes means that the corresponding variables do not influence each other directly. More formally, the qualitative part of a probabilistic network is an I-map of the represented probability distribution. Associated with the qualitative part G are numerical quantities from the joint probability distribution that is being represented. With each node A a set of conditional probability distributions is associated, describing the joint influence of the values of the nodes in $\pi_G(A)$ on the probabilities of the values of A . These probability distributions together constitute the quantitative part of the network.

We define the concept of a probabilistic network more formally.

Definition 2.26 A probabilistic network is a tuple $B = (G, \mathcal{P})$ where

- $G = (V(G), A(G))$ is an acyclic directed graph with nodes $V(G)$ and arcs $A(G)$;
- $\mathcal{P} = \{\Pr_A \mid A \in V(G)\}$, where, for each node $A \in V(G)$, $\Pr_A \in \mathcal{P}$ is a set of (conditional) probability distributions $\Pr(A \mid x)$ for each combination of values x for $\pi_G(A)$.

We illustrate the definition of a probabilistic network with an example.

Example 2.27 Consider the probabilistic network shown in Figure 2.1. The network represents

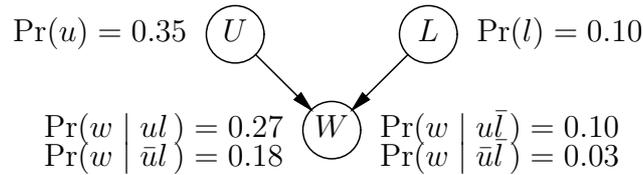


Figure 2.1: The *Wall Invasion* network.

a small, highly simplified fragment of the diagnostic part of the *oesophagus network*. Node U represents whether or not the carcinoma in a patient's oesophagus is ulcerating. Node L models the length of the carcinoma, where l denotes a length of 10 *cm* or more and \bar{l} is used to denote smaller lengths. An oesophageal carcinoma upon growth typically invades the oesophageal wall. The oesophageal wall consists of various layers; when the carcinoma has grown through all layers, it may invade neighbouring structures. The depth of invasion into the oesophageal wall is modelled by the node W , where w indicates that the carcinoma has grown beyond the oesophageal wall and is invading neighbouring structures; \bar{w} indicates that the invasion of the carcinoma is restricted to the oesophageal wall. Ulceration and the length of the carcinoma are modelled as the possible causes that influence the depth of invasion into the oesophageal wall. For the nodes U and L the network specifies the prior probability distributions $\Pr(U)$ and $\Pr(L)$, respectively. For node W it specifies four conditional probability distributions, one for each combination of values for the nodes U and L . These conditional distributions express, for example, that a small, non-ulcerating carcinoma is unlikely to have grown beyond the oesophageal wall to invade neighbouring structures. \square

A probabilistic network's qualitative and quantitative part together uniquely define a joint probability distribution that respects the independences portrayed by its digraph.

Proposition 2.28 Let $B = (G, \mathcal{P})$ be a probabilistic network with nodes $V(G)$. Then G is an *I-map* for the joint probability distribution \Pr on $V(G)$ defined by

$$\Pr(V(G)) = \prod_{A \in V(G)} \Pr(A \mid \pi_G(A)).$$

Since a probabilistic network uniquely defines a joint probability distribution, any prior or posterior probability of interest can be computed from the network. To this end, various efficient inference algorithms are available [73, 90]; it should be noted that exact probabilistic inference in probabilistic networks, based on the use of Bayes' rule to update probabilities, in general is NP-hard [22].

Qualitative Probabilistic Networks

Qualitative probabilistic networks were designed by M.P. Wellman as qualitative abstractions of probabilistic networks [134]. A qualitative probabilistic network bears a strong resemblance to its quantitative counterpart. It comprises a graphical representation of the independences holding among a set of variables, once more taking the form of an acyclic directed graph. Instead of conditional probabilities, however, a qualitative probabilistic network associates with its digraph *qualitative probabilistic relationships*. In Section 3.1 we formally define the concept of a qualitative probabilistic network; in Section 3.2 we will review an elegant algorithm for probabilistic inference with a qualitative network. While the discussion in Sections 3.1 and 3.2 focuses on binary nodes, Section 3.3 extends the discussion to non-binary nodes. A discussion of the advantages and disadvantages of the framework of qualitative probabilistic networks is provided in Section 3.4.

3.1 Defining a qualitative probabilistic network

A qualitative probabilistic network, as its quantitative counterpart, consists of an acyclic digraph $G = (V(G), A(G))$. The set of nodes $V(G)$ again represents the set of variables from the problem domain under consideration; we assume that for each variable a total order is specified on its values. The set of arcs $A(G)$ again models the independences holding between the variables, where independence is once more captured by the d-separation criterion. The digraph G can thus be considered an I-map of an existing, yet unknown joint probability distribution Pr . In addition to the digraph G , a qualitative probabilistic network, includes a set of hyperarcs for G . We recall that a probabilistic network, instead, includes a set of conditional probability distributions. The hyperarcs for G capture qualitative probabilistic relationships between variables, formally defined in terms of the probability distribution Pr . We distinguish between three types of qualitative probabilistic relationship: qualitative influences, additive synergies and product

synergies. In defining these relationships and their properties, we assume that none of the nodes in the digraph are observed; observation of nodes will be the subject of Section 3.2.

3.1.1 Qualitative influence

The most important type of qualitative relationship modelled in a qualitative probabilistic network is the qualitative influence. A qualitative influence between two nodes in the network's digraph G expresses how the values of one node influence the probabilities of the values of the other node. For example, a *positive qualitative influence* of a node A on its child B expresses that observing higher values for node A makes higher values for node B more likely, regardless of any other direct influences on B , where the concept of 'higher' refers to the order on a node's values. To express the probability that a node B has a value b_i or less, we introduce the concept of *cumulative probability*.

Definition 3.1 *Let U be a set of variables. Let \Pr be a joint probability distribution on U and let $B \in U$ with values $b_1 < \dots < b_n$, $n \geq 1$. Then, the function $F_B : \{b_1, \dots, b_n\} \rightarrow [0, 1]$ defined by*

$$F_B(b_i) = \Pr(b_1 \vee b_2 \vee \dots \vee b_i)$$

is the cumulative probability distribution function of B .

Given the total order on a node B 's values, instead of $\Pr(b_1 \vee b_2 \vee \dots \vee b_i)$ we will write $\Pr(B \leq b_i)$ for short. Cumulative *conditional* probability distribution functions are defined analogously.

Higher values for a node B are more likely given higher values for a node A , if the cumulative conditional probability distribution $F'_{B|a_i}$ of node B given a_i lies, graphically speaking, below the cumulative conditional probability distribution $F_{B|a_j}$ given a_j , for all values a_i, a_j of A with $a_i > a_j$. When $F'_{B|a_i}$ lies below $F_{B|a_j}$ for all values of B , $F'_{B|a_i}$ is said to dominate $F_{B|a_j}$ by *first-order stochastic dominance (FSD)*:

$$F'_{B|a_i} \text{ FSD } F_{B|a_j} \iff F'_{B|a_i}(b_i) \leq F_{B|a_j}(b_i) \text{ for all values } b_i \text{ of } B.$$

We illustrate the concept of dominance by means of an example.

Example 3.2 We consider a node B with values $b_1 < b_2 < b_3 < b_4$, and a node A with values a_i and a_j , $a_i > a_j$. The following conditional probability distributions are specified for B , given A :

$$\begin{array}{ll} \Pr(b_1 | a_i) = 0.25 & \Pr(b_1 | a_j) = 0.40 \\ \Pr(b_2 | a_i) = 0.35 & \Pr(b_2 | a_j) = 0.40 \\ \Pr(b_3 | a_i) = 0.25 & \Pr(b_3 | a_j) = 0.20 \\ \Pr(b_4 | a_i) = 0.15 & \Pr(b_4 | a_j) = 0.00 \end{array}$$

The *cumulative* conditional probability distributions for node B given A then are:

$$\begin{array}{ll} \Pr(B \leq b_1 | a_i) = 0.25 & \Pr(B \leq b_1 | a_j) = 0.40 \\ \Pr(B \leq b_2 | a_i) = 0.60 & \Pr(B \leq b_2 | a_j) = 0.80 \\ \Pr(B \leq b_3 | a_i) = 0.85 & \Pr(B \leq b_3 | a_j) = 1.00 \\ \Pr(B \leq b_4 | a_i) = 1.00 & \Pr(B \leq b_4 | a_j) = 1.00 \end{array}$$

We conclude that the cumulative conditional probability distribution for B given the value a_i of A dominates the cumulative conditional probability distribution for B given the value a_j . Indeed, we have from the conditional probability distributions specified for B that higher values for B are more likely for a_i than for a_j . \square

The concept of first-order stochastic dominance underlies the definition of qualitative influence.

Definition 3.3 *Let $G = (V(G), A(G))$ be an acyclic directed graph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let $A, B \in V(G)$ be nodes in G with $A \rightarrow B \in A(G)$. Let $X = \pi_G(B) \setminus \{A\}$. Then, node A positively influences node B along arc $A \rightarrow B$ in G , written $S_G^+(A, B)$, iff for all values b_i of B and all values a_j, a_k of A , with $a_j > a_k$, we have*

$$\Pr(B \geq b_i \mid a_j x) \geq \Pr(B \geq b_i \mid a_k x),$$

for any combination of values x for the set of nodes X .

For ease of exposition, we assume from here on that all nodes are binary-valued. Generalising the definitions we provide to non-binary nodes is straightforward using the definition above as an example; we will return to non-binary nodes in Section 3.3. A binary node A has the possible values *true* and *false*, with *true* $>$ *false*; as before, we will write a to denote the proposition $A = \text{true}$ and \bar{a} to denote $A = \text{false}$. For illustration purposes in examples, binary nodes often have different values than *true* and *false*; value statements for these nodes, however, are again written as a and \bar{a} , and we will equally assume $a > \bar{a}$. For binary nodes the definition of qualitative influence can be simplified.

Definition 3.4 *Let G, \Pr, A, B , and X be as in the previous definition. Then, node A positively influences node B along arc $A \rightarrow B$ in G iff*

$$\Pr(b \mid ax) \geq \Pr(b \mid \bar{a}x),$$

for any combination of values x for the set of nodes X .

The inequality from Definition 3.4 expresses that the influence of node A on node B along arc $A \rightarrow B$ is positive regardless of the probability distribution for the set of nodes X .

Lemma 3.5 *Let \Pr, A, B , and X be as before. If node A positively influences node B along arc $A \rightarrow B$, then*

$$\Pr(b \mid a) - \Pr(b \mid \bar{a}) \geq 0.$$

Proof: Suppose that we have

$$\Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \geq 0,$$

for all possible combinations of values x for the set of nodes X . As $\Pr(x) \geq 0$ for any combination of values x , we have that

$$\forall x \Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \geq 0 \implies \sum_x \Pr(x) \cdot (\Pr(b \mid ax) - \Pr(b \mid \bar{a}x)) \geq 0.$$

That is,

$$\forall x \Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \geq 0 \Rightarrow \left(\sum_x \Pr(x) \cdot \Pr(b \mid ax) \right) - \left(\sum_x \Pr(x) \cdot \Pr(b \mid \bar{a}x) \right) \geq 0.$$

Exploiting the fact that for the direct influence of node A on node B , node A can be regarded independent of the set of nodes X , the right-hand side of this implication equals

$$\left(\sum_x \Pr(b \mid ax) \cdot \Pr(x \mid a) \right) - \left(\sum_x \Pr(b \mid \bar{a}x) \cdot \Pr(x \mid \bar{a}) \right) \geq 0,$$

which is equivalent to $\Pr(b \mid a) - \Pr(b \mid \bar{a}) \geq 0$.

We therefore conclude that $\Pr(b \mid a) - \Pr(b \mid \bar{a}) \geq 0$. \square

The $+$ in the notation $S_G^+(A, B)$ for the qualitative influence defined above is called the *sign* of the influence of node A on node B . A *negative qualitative influence* of node A on its child B , denoted $S_G^-(A, B)$, and a *zero qualitative influence* of A on B , denoted $S_G^0(A, B)$, are defined analogously, replacing \geq in the above formula by \leq and $=$, respectively. Note that a zero influence of node A on node B indicates that A and B are (conditionally) independent. If the qualitative influence of A on B is *ambiguous*, that is, the influence is either *non-monotonic* or *unknown*, we write $S_G^?(A, B)$.

The following example illustrates the concept of qualitative influence by means of the example network from Section 2.4.

Example 3.6 We consider the qualitative abstraction of the *Wall Invasion* network from Figure 2.1. From the network the qualitative influences between the nodes are identified. For example, from the conditional probability distributions specified for node W we find that

$$\Pr(w \mid ul) \geq \Pr(w \mid \bar{u}l), \quad \text{and} \quad \Pr(w \mid u\bar{l}) \geq \Pr(w \mid \bar{u}\bar{l}).$$

So,

$$\Pr(w \mid ux) \geq \Pr(w \mid \bar{u}x),$$

for any value x of the other parents of node W , that is, other than U . We conclude that node U exerts a positive qualitative influence on node W : $S_G^+(U, W)$. It is easily verified that node L also exerts a positive qualitative influence on W . The qualitative probabilistic network that is thus abstracted from the *Wall Invasion* network is shown in Figure 3.1, depicting just the signs of

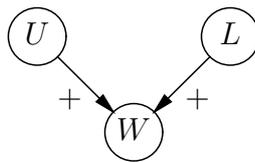


Figure 3.1: The direct qualitative influences of the qualitative *Wall Invasion* network.

the qualitative influences along the network's arcs. In this and following examples the qualitative influences between the nodes are computed from the probability distributions of the original, fully quantified probabilistic network. We would like to emphasise that in real-life applications of the framework of qualitative probabilistic networks these stochastic dominance statements are elicited directly from domain experts. \square

In a qualitative probabilistic network, a qualitative influence is associated with each arc of the network's digraph. Nodes, however, not only influence each other along arcs, they can also exert influences on one another indirectly. To capture *indirect* qualitative influences between two nodes, Wellman presents a set of reduction rules to collapse all trails between these nodes into a single arc, and to compute the sign of influence along this one arc from the signs associated with the collapsed arcs. To allow for reasoning about influences within the original digraph, that is, without having to reduce it, we take a different approach by defining a qualitative influence along a trail. To this end, we extend Definition 3.4.

Definition 3.7 *Let G , \Pr , A , and B be as in Definition 3.3. Let t be a trail from A to B in G . Let $X = (\bigcup_{C \in V(t) \setminus \{A\}} \pi_G(C)) \setminus V(t)$. Then, node A positively influences node B along trail t in G , written $\hat{S}_G^+(A, B, t)$, iff*

$$\Pr(b \mid ax) \geq \Pr(b \mid \bar{a}x),$$

for any combination of values x for the set of nodes X .

A *negative qualitative influence* of node A on node B along trail t , denoted $\hat{S}_G^-(A, B, t)$, and a *zero qualitative influence* of A on B along trail t , denoted $\hat{S}_G^0(A, B, t)$, are defined analogously, again replacing \geq in the above formula by \leq and $=$, respectively. If the qualitative influence of A on B along a trail t is *ambiguous*, we write $\hat{S}_G^?(A, B, t)$.

Properties of qualitative influences

The set of all direct and indirect qualitative influences between the nodes of a qualitative probabilistic network exhibits various convenient properties [99, 134]. We review the most important properties, formulating them in terms of trails; the properties are provided with simple proofs to serve as examples for the more complex proofs to follow. The properties we present allow us to determine the sign of a qualitative influence along a *sinkless* trail from the signs of the influences associated with the arcs on the trail. We will see in Section 3.2 that the properties thereby allow for an elegant sign-propagation algorithm for qualitative probabilistic inference.

Qualitative influences adhere to the property of *symmetry*. This property states that if a node A exerts a qualitative influence on a node B , then node B exerts a qualitative influence of the same sign on node A .

Proposition 3.8 *Let $G = (V(G), A(G))$ be an acyclic directed graph. Let $A, B \in V(G)$ be nodes in G . Let t be a trail from A to B in G , and let t^{-1} be its trail inverse. Then,*

$$\hat{S}_G^\delta(A, B, t) \iff \hat{S}_G^\delta(B, A, t^{-1}),$$

for any sign $\delta \in \{+, -, 0, ?\}$.

Proof: We prove the proposition for $\delta = +$; proofs for $\delta = -, 0$, and $?$ are analogous. Without loss of generality we now treat the sinkless trail t as if it were a single arc. We take the arc to be from node A to node B ; similar observations apply if we take the arc to be reversed. Let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let x be any combination of values for the set of nodes $X = \pi_G(B) \setminus \{A\}$. Then, by definition, we have

$$\begin{aligned}\hat{S}_G^+(A, B, t) &\iff \forall x \Pr(b \mid ax) \geq \Pr(b \mid \bar{a}x) \\ &\iff \forall x \Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \geq 0.\end{aligned}$$

Now, let y be any combination of values for the set of nodes $Y = \pi_G(A)$. We observe that, given A and X , node B is independent of Y . Using this property and multiplying by $\Pr(axy) \cdot \Pr(\bar{a}xy)$, we find

$$\begin{aligned}\hat{S}_G^+(A, B, t) &\iff \forall xy \Pr(b \mid axy) - \Pr(b \mid \bar{a}xy) \geq 0 \\ &\iff \forall xy \Pr(axy) \cdot \Pr(\bar{a}xy) - \Pr(\bar{a}bxy) \cdot \Pr(axy) \geq 0 \\ &\iff \forall xy \Pr(axy) \cdot (\Pr(\bar{a}bxy) + \Pr(\bar{a}\bar{b}xy)) \\ &\quad - \Pr(\bar{a}bxy) \cdot (\Pr(axy) + \Pr(\bar{a}\bar{b}xy)) \geq 0 \\ &\iff \forall xy \Pr(axy) \cdot \Pr(\bar{a}\bar{b}xy) - \Pr(\bar{a}bxy) \cdot \Pr(\bar{a}\bar{b}xy) \geq 0.\end{aligned}$$

By multiplying with $\frac{1}{\Pr(bxy) \cdot \Pr(\bar{b}xy)}$, we find

$$\begin{aligned}\hat{S}_G^+(A, B, t) &\iff \forall xy \frac{\Pr(axy) \cdot \Pr(\bar{a}\bar{b}xy)}{\Pr(bxy) \cdot \Pr(\bar{b}xy)} - \frac{\Pr(\bar{a}bxy) \cdot \Pr(\bar{a}\bar{b}xy)}{\Pr(bxy) \cdot \Pr(\bar{b}xy)} \geq 0 \\ &\iff \forall xy \Pr(a \mid bxy) \cdot \Pr(\bar{a} \mid \bar{b}xy) - \Pr(\bar{a} \mid bxy) \cdot \Pr(a \mid \bar{b}xy) \geq 0 \\ &\iff \forall xy \Pr(a \mid bxy) - \Pr(a \mid \bar{b}xy) \geq 0.\end{aligned}$$

We thus have

$$\hat{S}_G^+(A, B, t) \iff \hat{S}_G^+(B, A, t^{-1}),$$

by definition. □

We would like to note that qualitative influences are symmetric only with regard to their *sign*. The *strength* of a qualitative influence of a node A on a node B can differ considerably from the strength of the symmetric influence of B on A .

Qualitative influences in addition adhere to the property of *transitivity*. This property allows for the construction of the sign of a qualitative influence along a trail from the signs of the qualitative influences associated with its arcs. For example, we consider three nodes A, B and C with $A \rightarrow B$ and $B \rightarrow C$, and the trail t from A to C comprising these two arcs. The property of transitivity then states that the sign of the qualitative influence of node A on node C along trail t equals the ‘product’ of the signs of the qualitative influences associated with the two arcs on t . The ‘product’ operator \otimes , called the *sign-product*, for combining the signs is defined in Table 3.1. More generally, the property of transitivity applies to indirect qualitative influences along sinkless simple trails.

\otimes	+	-	0	?	\oplus	+	-	0	?
+	+	-	0	?	+	+	?	+	?
-	-	+	0	?	-	?	-	-	?
0	0	0	0	0	0	+	-	0	?
?	?	?	0	?	?	?	?	?	?

Table 3.1: The \otimes - and \oplus -operators for combining signs.

Proposition 3.9 *Let G be an acyclic digraph and let A, B, C be nodes in G . Let t_i and t_j be trails from A to B and from B to C in G , respectively, such that their trail concatenation $t_i \circ t_j$ is sinkless. Let \otimes be the operator defined in Table 3.1. Then,*

$$\hat{S}_G^{\delta_i}(A, B, t_i) \wedge \hat{S}_G^{\delta_j}(B, C, t_j) \implies \hat{S}_G^{\delta_i \otimes \delta_j}(A, C, t_i \circ t_j),$$

for any signs $\delta_i, \delta_j \in \{+, -, 0, ?\}$.

Proof: We prove the proposition for $\delta_i = \delta_j = +$; proofs for $\delta_i, \delta_j \in \{-, 0, ?\}$ are analogous. Note that $\delta_i \otimes \delta_j = + \otimes + = +$. Without loss of generality we now treat the (sinkless) trails t_i and t_j as if it were single arcs from A to B and B to C , respectively; similar observations apply with the arcs reversed, as long as the concatenation remains sinkless. Let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let x and y be any combination of values for the sets of nodes $X = \pi_G(C) \setminus \{B\}$ and $Y = \pi_G(B) \setminus \{A\}$, respectively. As we consider the trails t_i, t_j and their trail concatenation *in isolation*, we disregard any influence of node A on node B along trails other than t_i and t_j . As a result, the set of nodes X can be considered independent of node A , and node B is considered independent of X given Y . Proposition 3.10 will show how to determine the *net* influence of node A on node B along the parallel trail composition of all possible trails between A and B .

By definition, we have

$$\hat{S}_G^+(A, B, t_i) \wedge \hat{S}_G^+(B, C, t_j) \iff \forall y \Pr(b | ay) \geq \Pr(b | \bar{a}y) \wedge \forall x \Pr(c | bx) \geq \Pr(c | \bar{b}x),$$

which implies

$$(\Pr(b | ay) - \Pr(b | \bar{a}y)) \cdot (\Pr(c | bx) - \Pr(c | \bar{b}x)) \geq 0,$$

for any combination of values x and y for X and Y , respectively.

We now show, for any x and y ,

$$(\Pr(b | ay) - \Pr(b | \bar{a}y)) \cdot (\Pr(c | bx) - \Pr(c | \bar{b}x)) = \Pr(c | axy) - \Pr(c | \bar{a}xy).$$

By conditioning on node B , we find for any x and y

$$\begin{aligned} \Pr(c | axy) - \Pr(c | \bar{a}xy) &= \Pr(c | abxy) \cdot \Pr(b | axy) + \Pr(c | a\bar{b}xy) \cdot \Pr(\bar{b} | axy) \\ &\quad - \Pr(c | \bar{a}bxy) \cdot \Pr(b | \bar{a}xy) - \Pr(c | \bar{a}\bar{b}xy) \cdot \Pr(\bar{b} | \bar{a}xy) \\ &= \Pr(b | axy) \cdot (\Pr(c | abxy) - \Pr(c | \bar{a}bxy)) \\ &\quad - \Pr(b | \bar{a}xy) \cdot (\Pr(c | \bar{a}bxy) - \Pr(c | \bar{a}\bar{b}xy)) \\ &\quad + \Pr(c | a\bar{b}xy) - \Pr(c | \bar{a}\bar{b}xy). \end{aligned}$$

We observe that, given B and X , node C is independent of nodes $Y \cup \{A\}$; furthermore, node B is independent of X given Y . Using these properties, we find

$$\begin{aligned} \Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) &= \Pr(b \mid ay) \cdot (\Pr(c \mid bx) - \Pr(c \mid \bar{b}x)) + \Pr(c \mid \bar{b}x) \\ &\quad - \Pr(b \mid \bar{a}y) \cdot (\Pr(c \mid bx) - \Pr(c \mid \bar{b}x)) - \Pr(c \mid \bar{b}x) \\ &= (\Pr(b \mid ay) - \Pr(b \mid \bar{a}y)) \cdot (\Pr(c \mid bx) - \Pr(c \mid \bar{b}x)), \end{aligned}$$

for any x and y . We thus have

$$\begin{aligned} \hat{S}_G^+(A, B, t_i) \wedge \hat{S}_G^+(B, C, t_j) &\implies \forall xy \quad (\Pr(b \mid ay) - \Pr(b \mid \bar{a}y)) \cdot \\ &\quad \cdot (\Pr(c \mid bx) - \Pr(c \mid \bar{b}x)) \geq 0 \\ \iff \forall xy \quad \Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) &\geq 0 \\ \iff \hat{S}_G^+(A, B, t_i \circ t_j), & \end{aligned}$$

by definition. □

To conclude, qualitative influences adhere to the property of *composition*. This property states that the sign of the *net* influence of a node A on a node B along multiple parallel trails equals the ‘sum’ of the signs of the (indirect) qualitative influences of A on B along the various separate trails. The ‘sum’ operator \oplus , called the *sign-sum*, for combining the signs is defined in Table 3.1. The composition property allows for the construction of the sign of a qualitative influence along a composite trail from a node A to a node B from the signs of the qualitative influences along all sinkless simple trails from A to B constructed with the transitivity property.

Proposition 3.10 *Let G be an acyclic digraph and let A, B be nodes in G . Let t_i and t_j be trails from A to B in G such that node B is a head-to-head node on the trail concatenation $t_i \circ t_j^{-1}$. Let $t_i \parallel t_j$ be the parallel trail composition of the trails. Let \oplus be the operator defined in Table 3.1. Then,*

$$\hat{S}_G^{\delta_i}(A, B, t_i) \wedge \hat{S}_G^{\delta_j}(A, B, t_j) \implies \hat{S}_G^{\delta_i \oplus \delta_j}(A, B, t_i \parallel t_j),$$

for any signs $\delta_i, \delta_j \in \{+, -, 0, ?\}$.

Proof: We prove the proposition for $\delta_i = \delta_j = +$; proofs for $\delta_i, \delta_j \in \{-, 0, ?\}$ are analogous. Note that $\delta_i \oplus \delta_j = + \oplus + = +$. Without loss of generality we now treat the (sinkless) trail t_i as if it were a single arc from node A to node B and the (sinkless) trail t_j as if it consists of the arcs $A \rightarrow C, C \rightarrow B$ for some node $C \in V(G)$; similar observations apply with the arc between nodes A and C reversed. From $\hat{S}_G^+(A, B, t_j)$ and the property that trail t_j consists of the arcs $A \rightarrow C$ and $C \rightarrow B$, we deduce, using the property of transitivity and Table 3.1, that either

$$(1) \quad S_G^+(A, C) \wedge S_G^+(C, B), \quad \text{or} \quad (2) \quad S_G^-(A, C) \wedge S_G^-(C, B).$$

Suppose that situation (1) holds; similar observations hold for situation (2). Now, let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let x and y be any combination

of values for the sets of nodes $X = \pi_G(B) \setminus \{A, C\}$ and $Y = \pi_G(C) \setminus \{A\}$, respectively. Then, by definition, we have

$$\begin{aligned} \hat{S}_G^+(A, B, t_i) \wedge S_G^+(A, C) \wedge S_G^+(C, B) &\iff \forall x, y, a_i \in \{a, \bar{a}\}, c_i \in \{c, \bar{c}\} : \\ &\Pr(b \mid ac_ix) \geq \Pr(b \mid \bar{a}c_ix) \\ &\wedge \Pr(c \mid ay) \geq \Pr(c \mid \bar{a}y) \\ &\wedge \Pr(b \mid a_icx) \geq \Pr(b \mid a_i\bar{c}x). \end{aligned}$$

From this property, we now have to show that for any x and y

$$\Pr(b \mid axy) - \Pr(b \mid \bar{a}xy) \geq 0,$$

as this is, by definition, equivalent to $\hat{S}_G^+(A, B, t_i \parallel t_j)$.

We observe that, given A and C , node B is independent of Y ; furthermore, node C is independent of X given A . By conditioning on C and exploiting these independences, we find

$$\begin{aligned} \Pr(b \mid axy) - \Pr(b \mid \bar{a}xy) &= \Pr(b \mid acxy) \cdot \Pr(c \mid axy) + \Pr(b \mid \bar{a}cxy) \cdot \Pr(\bar{c} \mid axy) \\ &\quad - \Pr(b \mid \bar{a}cxy) \cdot \Pr(c \mid \bar{a}xy) - \Pr(b \mid \bar{a}\bar{c}xy) \cdot \Pr(\bar{c} \mid \bar{a}xy) \\ &= (\Pr(b \mid acx) - \Pr(b \mid \bar{a}cx)) \cdot \Pr(c \mid ay) + \Pr(b \mid \bar{a}cx) \\ &\quad - (\Pr(b \mid \bar{a}cx) - \Pr(b \mid \bar{a}\bar{c}x)) \cdot \Pr(c \mid \bar{a}y) - \Pr(b \mid \bar{a}\bar{c}x), \end{aligned}$$

for any combination of values x and y for X and Y , respectively. We know that for any x and y

$$\begin{aligned} \Pr(b \mid acx) \geq \Pr(b \mid \bar{a}cx), \quad \text{and} \quad \Pr(b \mid \bar{a}cx) \geq \Pr(b \mid \bar{a}\bar{c}x), \quad \text{and} \\ \Pr(c \mid ay) \geq \Pr(c \mid \bar{a}y), \quad \text{and} \quad \Pr(b \mid \bar{a}cx) \geq \Pr(b \mid \bar{a}\bar{c}x). \end{aligned}$$

These properties, however, do not suffice for determining the sign of

$$\begin{aligned} \Pr(b \mid axy) - \Pr(b \mid \bar{a}xy) &= (\Pr(b \mid acx) - \Pr(b \mid \bar{a}cx)) \cdot \Pr(c \mid ay) + \Pr(b \mid \bar{a}cx) \\ &\quad - (\Pr(b \mid \bar{a}cx) - \Pr(b \mid \bar{a}\bar{c}x)) \cdot \Pr(c \mid \bar{a}y) - \Pr(b \mid \bar{a}\bar{c}x). \end{aligned}$$

We therefore distinguish between the combinations of values x_i of X for which

$$\Pr(b \mid acx_i) - \Pr(b \mid \bar{a}cx_i) \geq \Pr(b \mid \bar{a}cx_i) - \Pr(b \mid \bar{a}\bar{c}x_i)$$

and the combinations of values x_i of X for which

$$\Pr(b \mid acx_i) - \Pr(b \mid \bar{a}cx_i) < \Pr(b \mid \bar{a}cx_i) - \Pr(b \mid \bar{a}\bar{c}x_i).$$

- Suppose that $\Pr(b \mid acx_i) - \Pr(b \mid \bar{a}cx_i) \geq \Pr(b \mid \bar{a}cx_i) - \Pr(b \mid \bar{a}\bar{c}x_i)$ for some combination of values x_i for X . We then find that $\Pr(b \mid ax_iy) - \Pr(b \mid \bar{a}x_iy) \geq 0$.
- Suppose that $\Pr(b \mid acx_i) - \Pr(b \mid \bar{a}cx_i) < \Pr(b \mid \bar{a}cx_i) - \Pr(b \mid \bar{a}\bar{c}x_i)$ for some combination of values x_i for X . Let p be short for $\Pr(b \mid ax_iy) - \Pr(b \mid \bar{a}x_iy)$. Note that p is a linear function in $\Pr(c \mid ay)$ and $\Pr(c \mid \bar{a}y)$. We will now show that it is not possible for p

to be negative. To this end, we consider for which values of $\Pr(c | ay)$ and $\Pr(c | \bar{a}y)$ the minimum of p is attained. The minimum of p is in principle attained for $\Pr(c | ay) = 0$ and $\Pr(c | \bar{a}y) = 1$. However, as node C exerts a positive influence on node B , we have the constraint that $\Pr(c | ay) \geq \Pr(c | \bar{a}y)$. We conclude that p 's minimum is attained for values of $\Pr(c | ay)$ and $\Pr(c | \bar{a}y)$ with $\Pr(c | ay) = \Pr(c | \bar{a}y)$. As p is a linear function, p is positive for any value of $\Pr(c | ay) = \Pr(c | \bar{a}y)$, if it is positive for the extreme values $\Pr(c | ay) = \Pr(c | \bar{a}y) = 0$ and $\Pr(c | ay) = \Pr(c | \bar{a}y) = 1$. It is easily verified that this is indeed the case.

Since we have made no assumptions about x_i , we conclude that

$$\Pr(b | axy) - \Pr(b | \bar{a}xy) \geq 0,$$

for any x and y , and therefore $\hat{S}_G^+(A, B, t_i) \wedge \hat{S}_G^+(A, B, t_j) \implies \hat{S}_G^+(A, B, t_i || t_j)$. \square

The signs of influences along trails are computed from the signs of the arcs on those trails using the symmetry, transitivity and composition properties. The conditions under which these properties hold ensure that the sign of an influence along a trail can only be computed for trails between nodes that are not d-separated from each other.

3.1.2 Additive synergy

In addition to qualitative influences, a qualitative probabilistic network includes *synergies*. A synergy models an interaction among the influences between three nodes in a network's digraph. We distinguish between two types of interaction, captured by additive and product synergies, respectively. In this section we focus on additive synergies; product synergies will be discussed in Section 3.1.3. An *additive synergy* expresses how the values of two nodes *jointly* influence the probabilities of the values of a third node [134]. For example, a *positive additive synergy* of nodes A and B on a common child C expresses that the joint influence of A and B on C is greater than the sum of their separate influences, regardless of any other direct influences on C .

Definition 3.11 *Let $G = (V(G), A(G))$ be an acyclic digraph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let $A, B, C \in V(G)$ be nodes in G with $A \rightarrow C, B \rightarrow C \in A(G)$. Let $X = \pi_G(C) \setminus \{A, B\}$. Then, nodes A and B exhibit a positive additive synergy on node C , written $Y_G^+(\{A, B\}, C)$, iff*

$$\Pr(c | abx) + \Pr(c | \bar{a}\bar{b}x) \geq \Pr(c | \bar{a}bx) + \Pr(c | a\bar{b}x),$$

for any combination of values x for the set of nodes X .

As with qualitative influences, if the inequality from Definition 3.11 holds for the nodes A , B , and C , then A and B exhibit a positive synergistic effect on C regardless of the probability distribution for the set of nodes X .

A *negative additive synergy* exhibited by nodes A and B on their common child C , denoted $Y_G^-(\{A, B\}, C)$, and a *zero additive synergy* of A and B on C , denoted $Y_G^0(\{A, B\}, C)$, are

defined analogously, once more replacing \geq in the above formula by \leq and $=$, respectively. If the additive synergy is *ambiguous*, we write $Y_G^?(\{A, B\}, C)$.

The following example illustrates the concept of additive synergy by means of the *Wall Invasion* probabilistic network.

Example 3.12 We consider once more the *Wall Invasion* network from Figure 2.1. We consider the additive synergies among the nodes in the network. From the conditional probability distributions specified for node W , we have for the joint influence of the various values for nodes U and L on W that

$$\begin{aligned}\Pr(w \mid ul) + \Pr(w \mid \bar{u}\bar{l}) &= 0.30, \\ \Pr(w \mid \bar{u}l) + \Pr(w \mid u\bar{l}) &= 0.28.\end{aligned}$$

So,

$$\Pr(w \mid ul) + \Pr(w \mid \bar{u}\bar{l}) \geq \Pr(w \mid \bar{u}l) + \Pr(w \mid u\bar{l}),$$

that is, the sum of the joint influences of the ‘same’ values for nodes U and L on W is *greater* than the sum of the joint influences of ‘different’ values for these nodes. We conclude that nodes U and L exhibit a positive additive synergy on node W , that is, $Y_G^+(\{U, L\}, W)$. \square

In a qualitative probabilistic network, nodes exhibit additive synergies on their common children. More generally, two nodes A and B can exhibit indirect additive synergies on a node C along trails from A to C and B to C , respectively. We briefly review the structural properties these trails have to adhere to. An additive synergy typically pertains to a head-to-head node. For an additive synergy of A and B on C along the trails t_i from A to C and t_j from B to C , we have that t_i and t_j should obey the following structural properties: t_i and t_j are sinkless and share a node $D \in V(t_i) \cap V(t_j)$ such that $t_i = t'_i \circ t_k$ and $t_j = t'_j \circ t_k$ share the subtrail t_k from D to C and D is a head-to-head node on all possible simple trails from A to B in $t'_i \circ t'_j^{-1}$. Note that when t_k is empty, node D is equivalent to node C ; otherwise node C is a descendant of D . The definition of additive synergy is now extended, analogous to the extension of the definition of qualitative influence, to capture additive synergies along such trails. From here on we will denote an additive synergy of sign δ exhibited by nodes A and B on a node C , along the trails t_i from A to C and t_j from B to C , respectively, by $\hat{Y}_G^\delta(\{A, B\}, C, \{t_i, t_j\})$.

Properties of additive synergies

The set of all direct and indirect additive synergies exhibits, just as the set of qualitative influences, various convenient properties. We briefly review these properties, formulating them in terms of trails; further details of the discussed properties can be found in [99, 134].

An additive synergy is trivially *symmetric*, in the sense that the two nodes that exhibit the synergy are interchangeable: $\hat{Y}_G^\delta(\{A, B\}, C, \{t_i, t_j\}) \iff \hat{Y}_G^\delta(\{B, A\}, C, \{t_j, t_i\})$. Additive synergies also adhere to a *transitivity* property, pertaining to the combination of an additive synergy with a qualitative influence. The property of transitivity states that the sign of the indirect additive synergy equals the sign-product of the signs of its ‘building blocks’, that is, the additive synergy and the qualitative influence. We distinguish between two types of transitivity: pre-synergy transitivity and post-synergy transitivity. In *pre-synergy* transitivity, an (indirect)

additive synergy exhibited by two nodes A and B on a node C is combined with a (indirect) qualitative influence from a fourth node D on either A or B , under the constraint that the resulting trail from D to C is sinkless. In *post-synergy* transitivity, an (indirect) additive synergy exhibited by A and B on C is combined with a (indirect) qualitative influence of C on a fourth node D , under the constraint that A and B are connected to D by sinkless trails that share the subtrail from C to D . Additive synergies further adhere to the property of *composition*. This property states that the sign of the *net* additive synergy of two nodes on a third node, along multiple parallel trails, equals the sign-sum of the signs of the separate additive synergies between these nodes.

3.1.3 Product synergy

In addition to additive synergies, a qualitative probabilistic network includes *product synergies* [56]. A product synergy expresses how the value of one node influences the probabilities of the values of another node upon knowing the value for a common child in the network's digraph. For example, a *negative product synergy* exhibited by nodes A and B with regard to the value c_0 for a common child C expresses that, given c_0 , observing higher values for node A renders higher values for node B less likely, regardless of any other influences on B , and vice versa. A product synergy thus describes the sign of the *intercausal* dependence between the two causes A and B that is induced by the observation of the value c_0 for the common effect C ; note that, given c_0 , the nodes A and B are no longer d-separated.

For formally defining the concept of product synergy, we address two nodes A and B with a common child C . We distinguish between two situations:

- Node C has no parents other than A and B , or all other parents are instantiated;
- Node C has uninstantiated parents other than A and B .

We begin by focusing on the situation where node C does not have any uninstantiated parents other than A and B . Nodes A and B then exhibit a product synergy of type I, or *product synergy I* for short, with regard to every single value for node C .

Definition 3.13 *Let $G = (V(G), A(G))$ be an acyclic directed graph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let $A, B, C \in V(G)$ be nodes in G with $A \rightarrow C, B \rightarrow C \in A(G)$. Let $X = \pi_G(C) \setminus \{A, B\}$ and let x_i be the combination of observed values for X . Then, nodes A and B exhibit a negative product synergy with regard to the value c_0 of node C , written $X_G^-(\{A, B\}, c_0)$, iff*

$$\Pr(c_0 \mid abx_i) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x_i) \leq \Pr(c_0 \mid \bar{a}bx_i) \cdot \Pr(c_0 \mid a\bar{b}x_i).$$

Positive, zero, and ambiguous product synergies once again are defined analogously.

We would like to note that additive and product synergies do not just differ in the addition and product-operators in their definitions. While an additive synergy exhibited by two nodes pertains to *all* values for a common child, a product synergy pertains to a *single* value for the child. There thus are as many product synergies as there are values for the child under consideration.

We illustrate the concept of product synergy by means of our running example.

Example 3.14 Again, we consider the *Wall Invasion* network from Figure 2.1. We focus on the product synergies among the nodes in the network. From the conditional probability distributions specified for node W we have that

$$\Pr(w \mid ul) \cdot \Pr(w \mid \bar{u}\bar{l}) = 0.0081, \quad \text{and} \quad \Pr(w \mid \bar{u}l) \cdot \Pr(w \mid u\bar{l}) = 0.018.$$

So,

$$\Pr(w \mid ul) \cdot \Pr(w \mid \bar{u}\bar{l}) \leq \Pr(w \mid \bar{u}l) \cdot \Pr(w \mid u\bar{l}).$$

We conclude that nodes U and L exhibit a negative product synergy with regard to the value w for node W . A similar observation holds with regard to the value \bar{w} for node W .

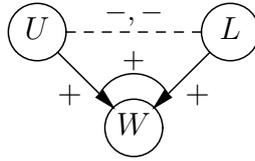


Figure 3.2: The qualitative *Wall Invasion* probabilistic network.

The qualitative probabilistic network that is thus abstracted from the *Wall Invasion* network is shown in Figure 3.2. The signs of the qualitative influences once again are shown along the network's arcs. The sign of the additive synergy exhibited by nodes U and L on node W , as found in Example 3.12, is indicated over the curve over node W . The signs of the product synergies exhibited by nodes U and L with regard to the different values for node W are indicated over the dashed edge between the associated nodes. \square

So far, we have considered product synergies of type I, exhibited by two nodes A and B with regard to the values for a common child C that has no uninstantiated parents other than A and B . We now focus on the more general situation where node C does have uninstantiated parents other than A and B . Let X be the set of parents of C . Recall that the signs of qualitative influences and additive synergies are independent of the probability distribution for the other nodes involved (see, for example, Lemma 3.5). If the definition of product synergy I would be adopted for the present situation, in contrast, it would not guarantee that the sign of the product synergy is independent of the distribution for X , that is, for example

$$\begin{aligned} \forall x \quad & \Pr(c_0 \mid abx) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x) - \Pr(c_0 \mid a\bar{b}x) \cdot \Pr(c_0 \mid \bar{a}bx) \leq 0 \\ \not\Rightarrow \quad & \Pr(c_0 \mid ab) \cdot \Pr(c_0 \mid \bar{a}\bar{b}) - \Pr(c_0 \mid a\bar{b}) \cdot \Pr(c_0 \mid \bar{a}b) \leq 0. \end{aligned}$$

To show this, we observe that

$$\begin{aligned} & \Pr(c_0 \mid ab) \cdot \Pr(c_0 \mid \bar{a}\bar{b}) - \Pr(c_0 \mid a\bar{b}) \cdot \Pr(c_0 \mid \bar{a}b) \\ &= \left(\sum_x \Pr(c \mid abx) \cdot \Pr(x) \right) \cdot \left(\sum_x \Pr(c \mid \bar{a}\bar{b}x) \cdot \Pr(x) \right) \end{aligned}$$

$$\begin{aligned}
& - \left(\sum_x \Pr(c \mid \bar{a}\bar{b}x) \cdot \Pr(x) \right) \cdot \left(\sum_x \Pr(c \mid \bar{a}bx) \cdot \Pr(x) \right) \\
& = \sum_{x_i} \sum_{x_j} \Pr(x_i) \cdot \Pr(x_j) \cdot (\Pr(c_0 \mid abx_i) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x_j) - \Pr(c_0 \mid \bar{a}\bar{b}x_i) \cdot \Pr(c_0 \mid \bar{a}bx_j)).
\end{aligned}$$

From

$$\forall x \Pr(c_0 \mid abx) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x) - \Pr(c_0 \mid \bar{a}\bar{b}x) \cdot \Pr(c_0 \mid \bar{a}bx) \leq 0$$

it does not follow, however, that

$$\sum_{x_i} \sum_{x_j} \Pr(x_i) \cdot \Pr(x_j) \cdot (\Pr(c_0 \mid abx_i) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x_j) - \Pr(c_0 \mid \bar{a}\bar{b}x_i) \cdot \Pr(c_0 \mid \bar{a}bx_j)) \leq 0.$$

In fact, we have that the sign of $p = \Pr(c_0 \mid ab) \cdot \Pr(c_0 \mid \bar{a}\bar{b}) - \Pr(c_0 \mid \bar{a}\bar{b}) \cdot \Pr(c_0 \mid \bar{a}b)$ may be dependent upon the probability distribution for the set X . More specifically, p is a *quadratic function* in the probability distribution for X . Now if X is observed, that is, if $\Pr(x) = 1$ for some combination of values x , then the sign of the product synergy equals the sign of the function p at the appropriate extreme. Product synergy I then correctly describes the sign of the synergy. If X is uninstantiated, however, then product synergy I no longer captures the correct sign of p : the function p may not have the same sign for every value of $\Pr(x)$, as is illustrated by Figure 3.3. M.J. Druzdzel and M. Henrion introduced an extended concept of product synergy, termed *product synergy II*, to capture the correct sign of p in the presence of uninstantiated parents [34]. Before formally defining product synergy II, we review the concept of *matrix half positive semi-definiteness*.

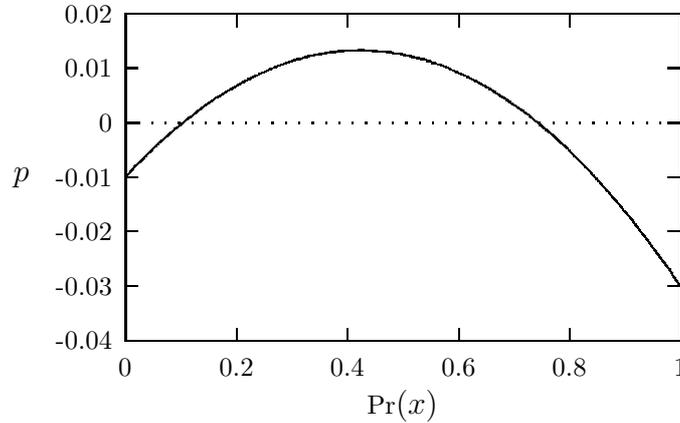


Figure 3.3: $p = \Pr(c_0 \mid ab) \cdot \Pr(c_0 \mid \bar{a}\bar{b}) - \Pr(c_0 \mid \bar{a}\bar{b}) \cdot \Pr(c_0 \mid \bar{a}b)$ as a function of the probability distribution of X .

Definition 3.15 Let M be a square $n \times n$ matrix, $n \geq 1$, and let x be any non-negative vector of n elements. Then, M is called half positive semi-definite if $x^T M x \geq 0$.

The concept of half negative semi-definiteness is defined analogously with \geq replaced by \leq . The following lemma states a useful property of a half positive semi-definite matrix.

Lemma 3.16 *If a square $n \times n$ matrix M , $n \geq 1$, is half positive semi-definite, then $M_{ii} \geq 0$ for $i = 1, \dots, n$.*

Proof: Let M be half positive semi-definite. Then, for any non-negative vector x of length n , we have $x^T M x \geq 0$. Now suppose there exists an $i \in \{1, \dots, n\}$ such that $M_{ii} < 0$. Let v be a non-negative vector that is non-zero only for v_i , then we have $v^T M v = v_i \cdot M_{ii} \cdot v_i < 0$, which contradicts the assumption that M is half positive semi-definite. \square

A similar property can be derived for a half negative semi-definite matrix. We now provide the definition of extended product synergy, termed *product synergy II*.

Definition 3.17 *Let $G = (V(G), A(G))$ be an acyclic digraph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let $A, B, C \in V(G)$ be nodes in G with $A \rightarrow C, B \rightarrow C \in A(G)$. Let $X = \pi_G(C) \setminus \{A, B\}$. Let m denote the number of possible combinations of values for X . Then, nodes A and B exhibit a negative product synergy with regard to the value c_0 for node C , iff the $m \times m$ -matrix D with elements*

$$D_{ij} = \Pr(c_0 \mid abx_i) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x_j) - \Pr(c_0 \mid \bar{a}bx_i) \cdot \Pr(c_0 \mid a\bar{b}x_j)$$

is half negative semi-definite for all combinations of values x_i and x_j for X .

For a positive or zero product synergy, the matrix D has to be half positive semi-definite or zero, respectively; if the matrix D does not exhibit any of these properties, the product synergy is ambiguous. Note that product synergy I is a special case of product synergy II. If the set of parents X , other than A and B , of node C is instantiated to a combination of values x , we can regard X as having only one possible value; the matrix D in Definition 3.17 then reduces to a matrix with a single element: $D_{11} = \Pr(c_0 \mid abx) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x) - \Pr(c_0 \mid \bar{a}bx) \cdot \Pr(c_0 \mid a\bar{b}x)$.

In a qualitative probabilistic network, nodes exhibit product synergies with regard to the values of their common children. More generally, two nodes A and B can exhibit indirect product synergies with regard to the values of a common descendant C . We briefly review the structural properties the trails involved have to adhere to. A product synergy typically pertains to the values of a head-to-head node or of one of its descendants. For a product synergy exhibited by A and B on the value c_0 of C along the trails t_i from A to C and t_j from B to C , we have that t_i and t_j should obey the following structural properties: C is a descendant of A and of B on t_i and on t_j , respectively, and t_i and t_j share a node $D \in V(t_i) \cap V(t_j)$ such that $t_i = t'_i \circ t_k$ and $t_j = t'_j \circ t_k$ share the subtrail t_k from D to C . Note again that when t_k is empty, node D is equivalent to node C . Also note that node D is a head-to-head node on all possible simple trails from A to B in $t'_i \circ t'_j^{-1}$. The definitions of product synergy are now extended, analogous to the extension of the definition of qualitative influence, to capture such indirect product synergies. From here on, we write $\hat{X}_G^\delta(\{A, B\}, c_0, \{t_i, t_j\})$ to denote a product synergy of sign δ exhibited by nodes A and B with regard to the value c_0 for a common descendant C along the trails t_i from A to C and t_j from B to C , respectively.

Properties of product synergies

The set of all product synergies exhibits, just as the sets of qualitative influences and additive synergies, various convenient properties. We will briefly review these properties here; further details can be found in [34, 99].

The properties for product synergies along trails closely resemble the properties for additive synergies along trails. The properties for product synergies once again include *symmetry*, *transitivity* and *composition* properties. A product synergy exhibited by nodes A and B with regard to their common descendant C is trivially symmetric in A and B . For the transitivity property, we again distinguish between *pre-synergy* and *post-synergy* transitivity. In pre-synergy transitivity a (indirect) product synergy exhibited by nodes A and B with regard to their common descendant C is combined with a (indirect) qualitative influence of a node D on either A or B , under the constraint that D is an ancestor of C on the resulting trail from D to C . In post-synergy transitivity, a (indirect) product synergy exhibited by nodes A and B with regard to their common descendant C is combined with a (indirect) qualitative influence of node C on a node D , under the constraint that D is a descendant of both A and B on the resulting trails. From the property of pre-synergy transitivity we have that the sign of the resulting indirect product synergy equals the sign-product of the signs of its building blocks, that is, the product synergy and the qualitative influence. For computing the sign of a product synergy resulting from post-synergy transitivity, the sign of the additive synergy between the nodes involved is also required; we refrain from further detailing this property here. Finally, product synergies adhere to the property of composition, which states that the sign of the *net* product synergy exhibited by two nodes with regard to the *same* value of a third node, along multiple parallel trails, equals the sign-sum of the signs of the separate product synergies among these nodes.

Intercausal reasoning

We mentioned before that the product synergy exhibited by two nodes A and B with regard to a specific value for a third node C , describes the dependence between A and B that is induced by the observation of the value for C . This dependence basically is an *influence* between A and B ; for this reason, we have indicated the signs of the product synergies over a dashed edge in Figure 3.2. We define the concept of intercausal influence more formally.

Definition 3.18 *Let $G = (V(G), A(G))$ be an acyclic digraph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let $A, B, C \in V(G)$ be nodes in G with $A \rightarrow C, B \rightarrow C \in A(G)$. Let $X = (\pi_G(B) \cup \pi_G(C)) \setminus \{A\}$. Then, given the value c_0 for node C , node A exhibits a positive intercausal influence on node B , written $Z_G^+(A, B, c_0)$, iff*

$$\Pr(b \mid ac_0x) \geq \Pr(b \mid \bar{a}c_0x)$$

for any combination of values x for the set of nodes X .

Negative, zero, and ambiguous intercausal influence again are defined analogously. From this definition, it is readily seen that an intercausal influence basically is a qualitative influence; it differs from a qualitative influence as defined before, only in that it describes an influence between two nodes along a trail that includes a head-to-head node. The definition of intercausal influence can be extended to intercausal influences along any trail t from a node A to a node B that includes a head-to-head node which is a common descendant of A and B , and which is instantiated or has an instantiated descendant; such an intercausal influence is denoted by $\hat{Z}_G^\delta(A, B, c_0, t)$ for the instantiation c_0 . The properties of symmetry, transitivity and parallel composition, as stated for qualitative influences, hold for intercausal influences as well.

The sign of the intercausal influence between A and B that is induced by the observation of a value for node C equals the sign of the product synergy with regard to that value for node C .

Proposition 3.19 *Let G , A , B , and C be as in the previous definition. Then, for each value c_0 for node C , we have*

$$X_G^\delta(\{A, B\}, c_0) \iff Z_G^\delta(A, B, c_0),$$

for any sign $\delta \in \{+, -, 0, ?\}$.

Proof: We prove the proposition for $\delta = +$; proofs for $\delta \in \{-, 0, ?\}$ are analogous. Let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let y be any combination of values for the set of nodes $Y = \pi_G(B) \setminus \{A\}$. To establish the sole effect of the intercausal influence, we assume there is no arc between nodes A and B . We prove the proposition first for product synergy I and then for product synergy II.

Let x be the instantiation of the set $X = \pi_G(C) \setminus \{A, B\}$. We have to prove that

$$\begin{aligned} & \Pr(c_0 \mid abx) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x) - \Pr(c_0 \mid \bar{a}bx) \cdot \Pr(c_0 \mid a\bar{b}x) \geq 0 \\ \iff & \forall y \Pr(b \mid ac_0xy) - \Pr(b \mid \bar{a}c_0xy) \geq 0. \end{aligned} \quad (3.1)$$

To this end, we will show that for any y

$$\begin{aligned} & \Pr(c_0 \mid abx) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x) - \Pr(c_0 \mid \bar{a}bx) \cdot \Pr(c_0 \mid a\bar{b}x) \\ &= \left(\Pr(b \mid ac_0xy) - \Pr(b \mid \bar{a}c_0xy) \right) \cdot \frac{\Pr(c_0 \mid axy) \cdot \Pr(c_0 \mid \bar{a}xy)}{\Pr(b \mid y) \cdot \Pr(\bar{b} \mid y)}. \end{aligned}$$

Using Bayes' theorem and exploiting independence, we find for an arbitrary value y_i of Y

$$\begin{aligned} & \Pr(c_0 \mid abx) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x) - \Pr(c_0 \mid \bar{a}bx) \cdot \Pr(c_0 \mid a\bar{b}x) \\ &= \Pr(c_0 \mid abxy_i) \cdot \Pr(c_0 \mid \bar{a}\bar{b}xy_i) - \Pr(c_0 \mid \bar{a}bxy_i) \cdot \Pr(c_0 \mid a\bar{b}xy_i) \\ &= \frac{\Pr(b \mid ac_0xy_i) \cdot \Pr(c_0 \mid axy_i)}{\Pr(b \mid y_i)} \cdot \frac{\Pr(\bar{b} \mid \bar{a}c_0xy_i) \cdot \Pr(c_0 \mid \bar{a}xy_i)}{\Pr(\bar{b} \mid y_i)} \\ & \quad - \frac{\Pr(b \mid \bar{a}c_0xy_i) \cdot \Pr(c_0 \mid \bar{a}xy_i)}{\Pr(b \mid y_i)} \cdot \frac{\Pr(\bar{b} \mid ac_0xy_i) \cdot \Pr(c_0 \mid axy_i)}{\Pr(\bar{b} \mid y_i)} \\ &= \frac{1}{\Pr(b \mid y_i) \cdot \Pr(\bar{b} \mid y_i)} \cdot \left(\Pr(b \mid ac_0xy_i) \cdot \Pr(c_0 \mid axy_i) \cdot \Pr(c_0 \mid \bar{a}xy_i) \right. \\ & \quad - \Pr(b \mid ac_0xy_i) \cdot \Pr(c_0 \mid axy_i) \cdot \Pr(b \mid \bar{a}c_0xy_i) \cdot \Pr(c_0 \mid \bar{a}xy_i) \\ & \quad - \Pr(b \mid \bar{a}c_0xy_i) \cdot \Pr(c_0 \mid \bar{a}xy_i) \cdot \Pr(c_0 \mid axy_i) \\ & \quad \left. + \Pr(b \mid \bar{a}c_0xy_i) \cdot \Pr(c_0 \mid \bar{a}xy_i) \cdot \Pr(b \mid ac_0xy_i) \cdot \Pr(c_0 \mid axy_i) \right) \\ &= \left(\Pr(b \mid ac_0xy_i) - \Pr(b \mid \bar{a}c_0xy_i) \right) \cdot \frac{\Pr(c_0 \mid axy_i) \cdot \Pr(c_0 \mid \bar{a}xy_i)}{\Pr(b \mid y_i) \cdot \Pr(\bar{b} \mid y_i)}. \end{aligned}$$

As we have made no assumptions about y_i , the above holds for any combination of values y of Y . From this equivalence we conclude that Equation 3.1 holds and hence that the proposition holds for product synergy as defined by product synergy I.

To prove the proposition for product synergy II, we assume, without loss of generality, that the set $X = \pi_G(C) \setminus \{A, B\}$ is uninstantiated. Let x , x_i and x_j be combinations of values for X . We now have to show that

the matrix D with $D_{ij} = \Pr(c_0 \mid abx_i) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x_j) - \Pr(c_0 \mid \bar{a}\bar{b}x_i) \cdot \Pr(c_0 \mid abx_j)$ is half positive semidefinite $\iff \forall xy \Pr(b \mid ac_0xy) - \Pr(b \mid \bar{a}c_0xy) \geq 0$.

We prove the two implications separately. Let the square matrix D be half positive semi-definite, then we have from Lemma 3.16 that

$$\forall i D_{ii} = \Pr(c_0 \mid abx_i) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x_i) - \Pr(c_0 \mid \bar{a}\bar{b}x_i) \cdot \Pr(c_0 \mid abx_i) \geq 0.$$

From Equation 3.1 above, we conclude that $\forall xy \Pr(b \mid ac_0xy) - \Pr(b \mid \bar{a}c_0xy) \geq 0$.

Now suppose that for all combinations of values x and y , we have

$$\Pr(b \mid ac_0xy) - \Pr(b \mid \bar{a}c_0xy) \geq 0.$$

From Lemma 3.5, we then find that $\Pr(b \mid ac_0) - \Pr(b \mid \bar{a}c_0) \geq 0$. Using Equation 3.1, we conclude that

$$\Pr(c_0 \mid ab) \cdot \Pr(c_0 \mid \bar{a}\bar{b}) - \Pr(c_0 \mid \bar{a}\bar{b}) \cdot \Pr(c_0 \mid ab) \geq 0.$$

From the discussion preceding Definition 3.15, we recall that

$$\Pr(c_0 \mid ab) \cdot \Pr(c_0 \mid \bar{a}\bar{b}) - \Pr(c_0 \mid \bar{a}\bar{b}) \cdot \Pr(c_0 \mid ab) \geq 0 \iff$$

$$\sum_{x_i} \sum_{x_j} \Pr(x_i) \cdot \Pr(x_j) \cdot \left(\Pr(c_0 \mid abx_i) \cdot \Pr(c_0 \mid \bar{a}\bar{b}x_j) - \Pr(c_0 \mid \bar{a}\bar{b}x_i) \cdot \Pr(c_0 \mid abx_j) \right) \geq 0,$$

from which we find $\sum_{x_i} \sum_{x_j} \Pr(x_i) \cdot \Pr(x_j) \cdot D_{ij} \geq 0$. This expression can be written as $p^T D p \geq$

0 for the non-negative vector p with $p_k = \Pr(x_k)$. Note that p is an arbitrary non-negative vector in de (probability) space we are considering. Therefore, D is half-positive semi-definite. We conclude that the proposition holds for product synergy II. \square

Intercausal influences, once induced, allow for reasoning about the dependences between the (indirect) causes of an observed common effect. This type of reasoning is termed intercausal reasoning and the most common pattern is known as *explaining away* [135]. Explaining away is evoked by a negative product synergy and describes the situation where one cause is sufficient to explain the observed effect: the other cause is then explained away. The idea of explaining away can be seen from the definition of product synergy. A negative product synergy exhibited by nodes A and B with regard to the value c_0 of their common descendant C can be written as

$$\frac{\Pr(c_0 \mid abx)}{\Pr(c_0 \mid \bar{a}\bar{b}x)} \leq \frac{\Pr(c_0 \mid \bar{a}\bar{b}x)}{\Pr(c_0 \mid abx)}.$$

The formula states that the proportional increase in the probability of c_0 upon switching the value of node B from *false* to *true*, is smaller in the context of a than in the context of \bar{a} ; the same holds for A and B reversed. We see, therefore, that the contribution of either cause being *true* to the probability of c_0 is greatest when it is the only cause present.

3.1.4 Definition of a qualitative probabilistic network

To conclude this section, we formally define the concept of a qualitative probabilistic network.

Definition 3.20 A qualitative probabilistic network is a tuple $Q = (G, \Delta)$, such that

- $G = (V(G), A(G))$ is an acyclic directed graph with nodes $V(G)$ and arcs $A(G)$;
- $\Delta = \mathcal{S} \cup \mathcal{Y} \cup \mathcal{X}$ is a set of hyperarcs for the digraph G where
 - \mathcal{S} is a set of qualitative influences for G such that
 - * \mathcal{S} includes a qualitative influence $S_G^\delta(A, B)$ for every two nodes $A, B \in V(G)$ with $A \rightarrow B \in A(G)$, and
 - * \mathcal{S} is closed under the properties of symmetry, transitivity, and composition;
 - \mathcal{Y} is a set of additive synergies for G such that
 - * \mathcal{Y} includes an additive synergy $Y_G^\delta(\{A, B\}, C)$ for every three nodes $A, B, C \in V(G)$ with $A \rightarrow C, B \rightarrow C \in A(G)$, and
 - * \mathcal{Y} is closed under the properties of symmetry, transitivity and composition;
 - \mathcal{X} is a set of product synergies for G such that
 - * \mathcal{X} includes a product synergy $X_G^\delta(\{A, B\}, c_0)$ for every three nodes $A, B, C \in V(G)$ with $A \rightarrow C, B \rightarrow C \in A(G)$ and every value c_0 for node C , and
 - * \mathcal{X} is closed under the properties of symmetry, transitivity and composition.

3.2 Inference in a qualitative probabilistic network

For probabilistic inference with a qualitative probabilistic network, an elegant algorithm is available, designed by M.J. Druzdzel and M. Henrion [33]. The basic idea of the algorithm is to trace the effect of observing a node's value upon the probabilities of the values of all other nodes in the network by message-passing between neighbouring nodes. In essence, this *sign-propagation* algorithm computes the sign of influence along the active trails between the observed node and all other nodes, using the properties of symmetry, transitivity and composition. All nodes that are not d-separated from the newly observed node by the set of all previously observed nodes end up with a *node sign* that indicates the direction of the shift in the node's probability distribution occasioned by the observation.

The sign-propagation algorithm is based on message-passing between neighbouring nodes. More specifically, a node sends messages to its currently *active neighbours*. We define the set of active neighbours for a node A that receives a message from a node B during inference.

Definition 3.21 Let $G = (V(G), A(G))$ be an acyclic digraph. Let $A, B \in V(G)$ be nodes in G such that, upon inference, node A receives a message from node B . Let $O \subseteq V(G)$ be the set of observed nodes in G . Let $X = \{X_i \mid X_i \in \sigma_G(A) \text{ and } \sigma_G^*(X_i) \cap O \neq \emptyset\}$ be the set of children of A with an observed descendant. Furthermore, let

$$N = \begin{cases} \sigma_G(A) \cup (\pi_G(X) \setminus \{A\}) & \text{if } B \rightarrow A \in A(G); \\ \pi_G(A) \cup (\sigma_G(A) \setminus \{B\}) \cup (\pi_G(X) \setminus \{A\}) & \text{if } A \rightarrow B \in A(G). \end{cases}$$

Then, an active neighbour of A is a node from $N \setminus O$.

Note that the set of active neighbours of a node A is dynamic as it depends on the node from which A receives a message during inference. We also note that node A 's current set of active neighbours is a subset of its Markov blanket: nodes from the set $\pi_G(\sigma_G(A))$ are only included if they are connected to A by an induced intercausal influence. For ease of exposition, we from here on assume that induced intercausal influences are added to the digraph as undirected edges.

procedure PropagateObservation($Q, O, sign, Observed$):

for each $V_i \in V(G)$
do sign[V_i] \leftarrow '0';
 PropagateSign($\emptyset, O, O, sign$).

procedure PropagateSign($trail, from, to, messagesign$):

sign[to] \leftarrow sign[to] \oplus messagesign;
 $trail \leftarrow trail \cup \{to\}$;
for each active neighbour V_i of to
do linksign \leftarrow sign of (induced) influence between to and V_i ;
 messagesign \leftarrow sign[to] \otimes linksign;
if $V_i \notin trail$ **and** sign[V_i] \neq sign[V_i] \oplus messagesign
then PropagateSign($trail, to, V_i, messagesign$).

Figure 3.4: The sign-propagation algorithm for probabilistic inference in a qualitative network.

The sign-propagation algorithm takes as input a qualitative probabilistic network Q , a set $Observed$ of previously observed nodes, a node O for which an observation has become available, and the sign $sign$ of the current observation, that is, either a '+' for the value *true* or a '-' for the value *false*. Prior to the propagation of the new observation, the node signs sign[V_i] for all nodes V_i are set to '0'. For the currently observed node O the appropriate sign is then entered into the network. The observed node updates its node sign to the sign-sum of its original sign and the entered sign. It thereupon notifies all its active neighbours that its sign has changed, by passing to each of them a message containing a sign; this sign is the sign-product of the node's current node sign and the sign *linksign* of the influence associated with the arc or edge it traverses. In each message it also records the origin of the message. A node to that receives a message, updates its node sign to the sign-sum of its current node sign sign[to] and the sign *messagesign* from the message it receives. The node then sends a copy of the message to all

its active neighbours that need to update their node sign. In doing so, the node changes the sign in each copy to the appropriate sign and adds itself as the origin of the copy. The information about the origin of the copy is added to the information about the origin of the message from which the copy resulted; as this process is repeated throughout the network, therefore, the trails along which messages have been passed are recorded. As messages travel simple trails only, it is sufficient to just record the nodes on these trails. The information is exploited in preventing the passing of messages to nodes that were already visited on the same trail.

During sign-propagation, nodes are only visited if they need a change of node sign. A node sign can change at most twice, once from '0' to '+', '-' or '?' and then only from '+' or '-' to '?'. From this observation we have that no node is ever visited more than twice upon inference and the algorithm is therefore guaranteed to halt. The time-complexity of the algorithm is linear in the number of arcs of the digraph. For a proof of the algorithm's correctness we refer the reader to [31]. The sign-propagation algorithm for probabilistic inference with a qualitative network is summarised in pseudocode in Figure 3.4; the basic idea of the algorithm is illustrated with an example.

Example 3.22 We consider the qualitative probabilistic network shown in Figure 3.5(a). The qualitative influences between the nodes in the network are indicated over the digraph's arcs, as before. We assume that the value of node D was observed to be *true* and that it has induced a positive intercausal influence between the nodes C and E , indicated in the figure by the dashed edge between these nodes. Now suppose that we are interested in the effect of observing the value *true* for node C upon all the other nodes in the network, in the presence of the previous observation for node D . Prior to the inference, the node signs for all nodes are set to '0', as depicted in the figure.

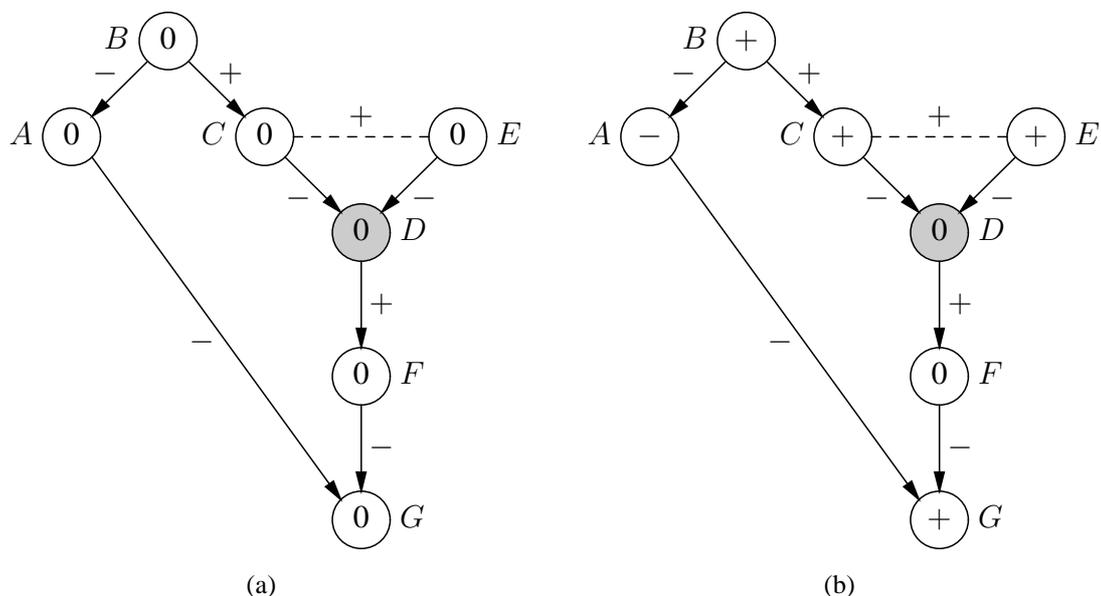


Figure 3.5: An example qualitative probabilistic network with the node signs before, (a), and after, (b), probabilistic inference.

To enter the observation for node C , the sign ‘+’ is entered into the network. Node C updates its node sign to the sign $0 \oplus + = +$ and subsequently determines the proper messages to be sent to its active neighbours B and E . No message will be sent to node D as it is observed and therefore d-separated from node C . For node B , node C computes the sign $+ \otimes + = +$. It thereupon sends a message containing the computed ‘+’ and the trail $[C]$ to node B . Note that in this step of the algorithm, the property of symmetry of influences is used explicitly. Upon receiving the message with the sign ‘+’, node B updates its node sign to $0 \oplus + = +$. It subsequently computes the sign $+ \otimes - = -$ to be sent to node A ; in the message also the trail $[C, B]$ is recorded. Note that for this purpose, the algorithm exploits the transitivity property of influences. After updating its node sign to $0 \oplus - = -$, node A sends a message with the sign $- \otimes - = +$ and the trail $[C, B, A]$ to node G , causing it to update its sign to $0 \oplus + = +$. No copy of the message originating from node A is passed on by node G to node F as node F is not an active neighbour of node G . For its induced neighbour E , node C computes the sign $+ \oplus + = +$. Node C sends a message containing this sign and the trail $[C]$ to E , causing it to update its node sign to ‘+’. Node E does not send any message to node D as this node has been observed. The process of message passing now halts. The result of the inference is depicted in Figure 3.5(b), showing the various node signs.

Note that if node D had not been previously observed, then there would exist two trails between node C and node G . If, in addition, the influence associated with the arc $C \rightarrow D$ were positive, the two trails would have conflicting signs and entering an observation for node C would then result in the node sign ‘?’ for node G . \square

3.3 A note on non-binary nodes

In Section 3.1, we provided the general definition of a qualitative influence and its simplification to binary nodes. For ease of exposition, we from there on assumed all nodes to be binary. Since all definitions and propositions presented in Section 3.1 can be generalised to non-binary nodes, our assumption is not a restrictive one. In this section, we will argue that a generalisation to non-binary nodes does not enhance results upon inference with the basic sign-propagation algorithm, as the generalisation does not provide us with a finer level of detail.

We recall from Section 3.2 that in the sign-propagation algorithm, an observation is entered as a ‘+’ or as a ‘-’. Although it is clear what these signs mean for a binary node, the meaning for a non-binary node is not so obvious. For a non-binary node A , a ‘+’ should be entered if ‘a higher value’ for A is observed. Not only does this assume a total order on the values for A , it also assumes that we can distinguish between ‘higher’ and ‘lower’ values. To this end, we will now assume that each node has a *dummy value* representing an initial ‘medium’ state in case we do not know the node’s actual value. We illustrate the use of this dummy value with an example.

Example 3.23 We consider the small part of the oesophagus network shown in Figure 3.6; it describes the extent of the lymph node metastases of an oesophageal carcinoma as an effect of the depth of invasion into the oesophageal wall. Node W , modelling the depth of invasion, has the four values $T_1 < T_2 < T_3 < T_4$, where T_1 denotes that the carcinoma has invaded only the first layer of the oesophageal wall and T_4 models that the carcinoma has grown beyond the

oesophageal wall. Node L , modelling the extent of the carcinoma's lymph node metastases, has the three values $N_0 < N_1 < M_1$, where N_0 indicates that there are no lymph node metastases and M_1 indicates the presence of metastases in distant lymph nodes. From domain experts we know

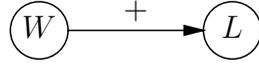


Figure 3.6: The *Wall invasion* network.

that T_1 and T_2 are considered lower values for the depth of invasion and T_3 and T_4 are regarded as higher values. The dummy value T_d , which we position between T_2 and T_3 in the ordering, is now used to denote the ‘medium’ value for node W . An observation of either T_3 or T_4 for node W is an observation of a value larger than the dummy value T_d ; therefore a ‘+’ is entered into the network. Similarly, a ‘-’ is entered into the network for the observation of either T_1 or T_2 . Entering a ‘+’ will then cause the values of node L that are higher than its dummy value to become more likely during inference. \square

The position of the dummy value for a node is defined by domain experts. To this end, experts are asked to partition the set of values of a node into two subsets: one subset with ‘high’ values and one with ‘low’ values. For an arc between two nodes A and B , the partition of the set of node A 's values can, however, depend on A 's influence on node B , and vice-versa. For example, the depth of wall invasion in the previous example not only influences the extent of lymph node metastases, but also the extent of haematogeneous metastases. For the influence of node W on lymph node metastases, T_3 and T_4 are considered ‘higher’ values for W ; haematogeneous metastases, however, only occur with a depth of invasion T_4 , which means that T_4 is then considered the only ‘higher’ value for node W . We can also illustrate this by formally defining, for an arc between a node A and a node B , the positions of the dummy values in the orderings of node A 's and node B 's values. Let nodes A and B have the values $a_1 < \dots < a_n$, $n \geq 1$, and $b_1 < \dots < b_m$, $m \geq 1$, respectively. Let a_d , b_d be the dummy values for node A and node B , respectively. Suppose node A exerts a positive qualitative influence on node B ; similar observations apply when the qualitative influence of A on B is negative. From the information that the influence of A on B is positive, we observe that

$$\Pr(B \geq b_k \mid A > a_d) \geq \Pr(B \geq b_k) \text{ and } \Pr(B \geq b_k \mid A < a_d) \leq \Pr(B \geq b_k)$$

should hold for all $k \in \{1, \dots, m\}$. This property can be attained by defining the probabilities for the separate values of node B given the dummy value a_d of node A to be equal to the prior probability of the value:

$$\Pr(B = b_k) = \Pr(B = b_k \mid a_d).$$

For each value b_k of node B we can now construct a graph similar to the one shown in Figure 3.7; note that $\Pr(B \geq b_k \mid a_i)$ is ascending for higher values a_i because A exerts a positive qualitative influence on B . Now suppose that for a given value b_k of node B , we know that

$$\Pr(B \geq b_k) = \sum_i \Pr(B \geq b_k \mid a_i) \cdot \Pr(a_i) \text{ equals } p, \text{ or equivalently, } \Pr(B \geq b_k \mid a_d) = p.$$

From the graph we can then easily determine the position of the dummy value a_d in the ordering of A 's values. Note that the dummy value a_d does not bear any relation with the a priori distribution for node A . Exploiting the property of symmetry of qualitative influences, the position of the dummy value b_d in the ordering of node B 's values can be determined in a similar way.

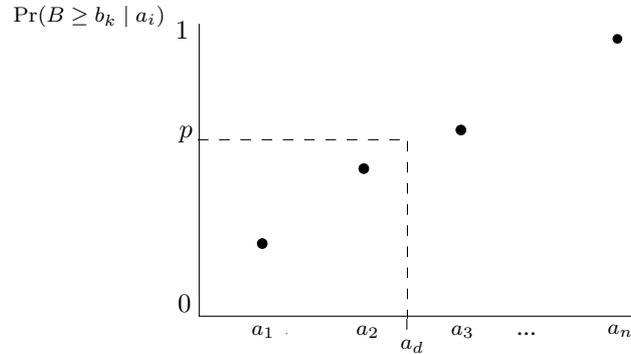


Figure 3.7: A graph of the cumulative probabilities for a value b_k of node B , given values of node A .

Unfortunately, the position of a node's dummy value need not be unique. For example, we consider a node A that is connected by two arcs to both node B and node C . It is possible that the position of the dummy value a_d for node A determined from node B 's cumulative probabilities differs from the position of the dummy value a'_d determined from node C 's cumulative probabilities. Now, suppose that node A exerts a positive qualitative influence on both nodes B and C , and suppose that $a_d > a'_d$. Then, only observations for values of node A larger than a_d will cause higher values of both B and C to become more likely; similarly, higher values of both B and C are ensured to become less likely only for observations of node A smaller than a'_d . Consequently, any observation for node A that lies between a'_d and a_d cannot be entered into the basic sign-propagation algorithm as a '+' or '-' and should be entered as a '?'. We would like to note that an observation between a'_d and a_d should in fact be propagated as a '-' to node B and as a '+' to node C . The basic sign-propagation algorithm must be adapted to enable the propagation of two different signs for a single observation.

Using the concept of dummy value it is now possible to determine the sign of an observation for a non-binary node and to enter it in a qualitative probabilistic network. The observation is then propagated using the sign-propagation algorithm. During the propagation, however, it is still only determined whether the observation has a positive or a negative influence on the different nodes in the network. As a result, sign-propagation with non-binary nodes provides us with no more information than sign-propagation using binary nodes only. We conclude that with the current level of detail used in the inference algorithm, a qualitative network including only binary nodes is just as expressive as one including non-binary nodes. We would like to add to this observation that for the purpose of quantification, a qualitative probabilistic network modelling non-binary nodes may nonetheless be preferred, as it may be a more realistic representation of the problem domain at hand.

3.4 Discussion

Qualitative probabilistic networks allow for modelling, in a simple and intuitively appealing way, the qualitative relationships between two nodes, as well as the interactions among more than two nodes. The probabilistic definitions of qualitative influences and synergies allow for probabilistic inference in an elegant and mathematically correct way, without having to elicit a large number of probabilities from domain experts. Despite these advantages, the formalism of qualitative probabilistic networks is also rather limited.

Qualitative probabilistic networks model the influences between their nodes at a coarse level of detail: influences are either positive, negative, zero, or ambiguous, without an indication of their strengths. Although this level of detail will suffice for some domains of application, there are problems that require a finer level of detail for their solution. One of the major drawbacks of the coarse level of detail in qualitative networks is the ease with which the uninformative ‘?’-sign arises: once an ambiguous sign is generated during inference, it will spread throughout most of the network. We can identify two causes for ambiguous signs arising during inference: the network includes an arc with an associated non-monotonic influence, or the network models one or more trade-offs. An influence of a node A on a node B is called *non-monotonic* if its sign depends upon the value of some third node; for non-binary nodes, the order of their values can also cause non-monotonicity of influences. As non-monotonicity cannot be modelled explicitly in a qualitative network, these influences are captured by the ‘?’-sign. A network models a *trade-off* if two nodes in the network are connected by multiple parallel trails and the signs of the influences along these trails are conflicting; conflicting influences between two nodes can also result from induced intercausal influences. As the high level of abstraction does not provide for a notion of strength of influences and, hence does not provide for weighing conflicting influences, the net influence between the two nodes is then unknown and a ‘?’ arises.

In the next chapter, we will address the problem of ambiguous signs more thoroughly. We will extend the formalism of qualitative probabilistic networks to enable the modelling of additional qualitative information. In addition, we will provide algorithms for handling and, if possible, preventing ambiguous signs. We feel that it is worthwhile to address this drawback of qualitative probabilistic networks and make the formalism as expressive as possible, because we believe that qualitative networks can play an important role in the construction of quantitative probabilistic networks for real-life application domains. Before assessing all conditional probabilities required for a probabilistic network, the qualitative probabilistic relationships can be elicited. As these concern stochastic dominance statements, they are more easily provided by domain experts than the probabilities required. We can then study the reasoning behaviour of the network under construction; this allows us to validate, at least to some extent, the network’s structure. When the network’s graphical structure is considered robust, the qualitative probabilistic relationships provide constraints on the required probability distributions that can be used as a guideline for assessing the numerical quantities.

Refining Qualitative Networks

One of the major drawbacks of qualitative probabilistic networks is their coarse level of representation detail. In a qualitative probabilistic network, the influence between two nodes is quantified with a sign; this sign is independent of any other influences in the network and has no indication of the strength of the influence. In Section 3.4 of Chapter 3, we discussed the consequences of this coarse level of representation detail. For example, a non-monotonic influence cannot be distinguished from an unknown influence as its sign depends on the value of some third node, and a trade-off cannot be resolved during inference because modelling the intricacies involved in weighing its conflicting influences is not possible. As we have argued before, probabilistic inference with a qualitative network can thus easily result in ambiguous node signs. Ambiguous results can be averted by refining the formalism of qualitative probabilistic networks to provide for a finer level of detail. Roughly speaking, the finer the level of detail, the more ambiguous results can be prevented. However, the finer the level of detail, the closer the network will resemble a fully quantified network, with the risk of losing the computational efficiency of qualitative reasoning.

In this chapter we propose several extensions to the framework of qualitative probabilistic networks that provide for stronger results upon inference, yet retain computational efficiency. Throughout the chapter we assume that additional qualitative properties, resulting from the proposed extensions, are added to the set Δ of qualitative properties of a qualitative probabilistic network. In Section 4.1 we propose a refinement of the basic formalism of qualitative probabilistic networks that concerns non-monotonic influences. We propose to distinguish between the representation of non-monotonic and unknown influences. In addition, we present a method for resolving the non-monotonicity. In Sections 4.2 and 4.3 we propose two well-defined, purely qualitative approaches to handling trade-offs. The first approach is an enhancement of the basic formalism of qualitative probabilistic networks that includes a notion of strength. This is achieved by extending the formalism with double signs $++$ and $--$; these double signs are

taken to outweigh conflicting $+$ and $-$ signs and can therefore help to resolve trade-offs. The second approach allows us to include context-specific information with each sign. In Section 4.4 we propose an algorithm for isolating, from a network, the trade-offs that cannot be resolved with the previously discussed refinements. If, during inference, an ambiguous node sign results for some node of interest, the algorithm traces the origin of the ambiguous sign and determines the additional information that would be required to resolve it. In Section 4.5 we extend the sign-propagation algorithm to the propagation of multiple simultaneous observations. Contrary to earlier proposals for handling multiple observations, our extension leads to the strongest possible results. We conclude the chapter with a brief overview of related work in Section 4.6.

4.1 Exploiting non-monotonic influences

A qualitative probabilistic network basically models *monotonic* qualitative influences between its nodes only. We recall from Chapter 3 that a qualitative influence exerted by a node A on a node B results in a shift in the probabilities of B 's values in a direction that is independent of any other influences exerted on B . Qualitative influences between nodes, however, need not necessarily be monotonic in nature: we say that the influence exerted by A on B is *non-monotonic* if the resulting direction of shift in the probabilities of B 's values *does* depend upon the influences of one or more other nodes.

In a qualitative probabilistic network, a non-monotonic qualitative influence between two nodes is captured by a '?'. The same sign is used to express an *unknown* qualitative influence, that is, a probabilistic influence for which the direction of shift is unknown. An unknown qualitative influence can be viewed as expressing lack of information: there is no information present in the network to conclude whether the sign of the influence is '+', '-', or '0'. A non-monotonic influence, on the other hand, conveys at least some information by the nature of its non-monotonicity. A non-monotonic influence can in fact be seen as expressing incomplete knowledge: the sign of the influence is as yet unknown, but it will be known when the values of some other nodes are observed. Non-monotonicity of a qualitative influence and lack of information, although expressed in the same way, are therefore different from a conceptual point of view. In this section we argue that it is worthwhile to explicitly distinguish between non-monotonic influences and unknown influences in a qualitative probabilistic network. We show how information that can be extracted from the non-monotonic influences of a network, can be exploited during probabilistic inference to prevent unnecessarily weak results.

4.1.1 Non-monotonic influences

A non-monotonic qualitative influence of a node A on a node B is an influence whose sign depends on the influences of other nodes on B .

Definition 4.1 *Let $G = (V(G), A(G))$ be an acyclic digraph and let Pr be a joint probability distribution on $V(G)$ such that G is an I-map for Pr . Let A and B be nodes in G with $A \rightarrow B \in A(G)$. Let $Y = \pi_G(B) \setminus \{A\}$. Then, the qualitative influence of node A on node B in G is*

non-monotonic, *iff there exist combinations of values y and y' for Y , such that*

$$\Pr(b | ay) > \Pr(b | \bar{a}y) \text{ and } \Pr(b | ay') < \Pr(b | \bar{a}y').$$

For ease of exposition, we discuss non-monotonic influences associated with arcs only. Generalisation to trails between nodes is straightforward.

Figure 4.1 illustrates the concept of a non-monotonic influence for three nodes A , B and C with $A \rightarrow B$ and $C \rightarrow B$. We observe that node C exerts a monotonic qualitative influence of sign ‘-’ on node B , for we have

$$\Pr(b | a_i c) \leq \Pr(b | a_i \bar{c}),$$

for each value $a_i \in \{a, \bar{a}\}$ for node A . The influence of A on node B , however, is non-monotonic, since the sign of the difference

$$\Pr(b | ac_i) - \Pr(b | \bar{a}c_i)$$

depends on the value $c_i \in \{c, \bar{c}\}$ for C . More specifically, we observe from the figure that

$$\Pr(b | ac) > \Pr(b | \bar{a}c) \text{ and } \Pr(b | a\bar{c}) < \Pr(b | \bar{a}\bar{c}).$$

Note that the qualitative influence of node A on node B becomes monotonic, as soon as an observation for node C is available.

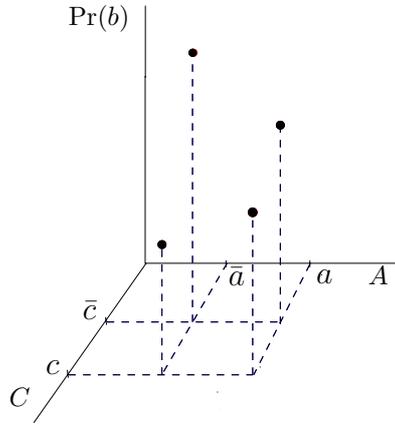


Figure 4.1: A non-monotonic influence of node A on node B provoked by node C .

As mentioned before, both non-monotonic influences and unknown influences are denoted by the sign ‘?’ in a qualitative probabilistic network. Now, in Chapter 3 we saw that the sign ‘?’ gives rise to ambiguous results during probabilistic inference in a qualitative network. It is therefore worthwhile to deal with the ‘?’-signs that are specified in the network by extracting as much information from them as possible. To this end, we explicitly distinguish between non-monotonic and unknown qualitative influences, by capturing a non-monotonic influence by the new sign ‘ \sim ’.

To deal with a non-monotonic qualitative influence of a node A on a node B , we observe that the influence is not positive, negative, zero, or unknown, but that its sign depends on the

value of one or more other nodes. We use the term *provokers* to denote the nodes on whose combination of values an influence's signs depends; we say that the non-monotonicity of the influence is *provoked* by this set of nodes. The fact that the non-monotonicity of an influence is provoked by a set of nodes P is indicated with the augmented sign ' \sim_P '. We will subsequently discuss non-monotonic influences with a single provoker, with two provokers, and with larger provoking sets.

One provoker

We formally define the concept of a single provoker of a non-monotonicity.

Definition 4.2 Let $G = (V(G), A(G))$ be an acyclic digraph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let A, B, C be nodes in G with $A \rightarrow B, C \rightarrow B \in A(G)$ such that A exerts a non-monotonic influence on B . Let $Y = \pi_G(B) \setminus \{A, C\}$. Then, node C is the provoker of the non-monotonicity of the influence of node A on node B in G , iff for some $c_i, c_j \in \{c, \bar{c}\}$, $c_i \neq c_j$, we have that

$$\Pr(b \mid ac_i y) \geq \Pr(b \mid \bar{a}c_i y) \text{ and } \Pr(b \mid ac_j y) \leq \Pr(b \mid \bar{a}c_j y),$$

for any combination of values y for Y .

From the previous definition we have that for each $c_k \in \{c, \bar{c}\}$, either $\Pr(b \mid ac_k y) \geq \Pr(b \mid \bar{a}c_k y)$ for any combination of values y for Y , or $\Pr(b \mid ac_k y) \leq \Pr(b \mid \bar{a}c_k y)$ for any such y .

To avoid an abundance of braces, we will write ' \sim_C ' instead of ' $\sim_{\{C\}}$ ' to indicate the sign of a non-monotonic influence with a single provoker C . The concept of a non-monotonicity that is provoked by a single node is illustrated with an example.

Example 4.3 We consider the probabilistic *Cervical Metastases* network and its qualitative abstraction shown in Figure 4.2. The network represents a small, highly simplified fragment of the

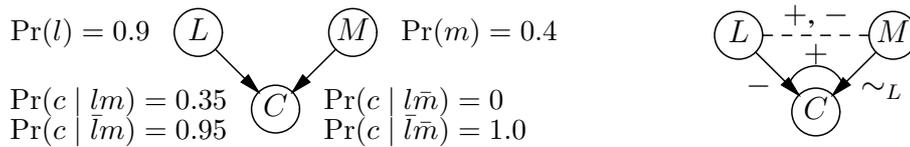


Figure 4.2: The *Cervical Metastases* network.

oesophagus network, pertaining to lymphatic metastases of a carcinoma. The node L represents the location of an oesophageal carcinoma in a patient's oesophagus. The fact that the tumour resides in the lower two-third of the oesophagus is represented by l ; \bar{l} expresses that the tumour is located in the oesophagus' upper one-third. An oesophageal carcinoma upon growth typically gives rise to lymphatic metastases. Node M represents the extent of these metastases. If the distant lymph nodes are affected by cancer cells this is indicated by m ; \bar{m} denotes that just the local and regional lymph nodes are affected. Which lymph nodes are local or regional and which are distant depends on the location of the primary tumour in the oesophagus. The lymph nodes in the

neck, or cervix, for example, are regional for a tumour in the upper one-third of the oesophagus and distant otherwise. Node C represents the presence or absence in a patient of metastases in the cervical lymph nodes.

From the conditional probabilities specified for node C , it is readily verified that node L exerts a negative qualitative influence on C . The influence of node M on C is non-monotonic. The non-monotonicity of the influence is provoked by node L , the location of the carcinoma. Furthermore, we observe that nodes L and M exhibit a positive additive synergy on C and either value for the node C induces an intercausal influence between L and M . For the value c this intercausal influence is captured by a positive product synergy and for the value \bar{c} the influence is captured by a negative synergy. \square

From the definition of a single provoker C of the non-monotonicity of the influence of node A on node B , we have more specifically that for all combinations of values y for the set Y of parents of B other than A and C , either

$$\Pr(b \mid acy) \geq \Pr(b \mid \bar{a}cy) \quad \text{and} \quad \Pr(b \mid a\bar{c}y) \leq \Pr(b \mid \bar{a}\bar{c}y),$$

or

$$\Pr(b \mid acy) \leq \Pr(b \mid \bar{a}cy) \quad \text{and} \quad \Pr(b \mid a\bar{c}y) \geq \Pr(b \mid \bar{a}\bar{c}y),$$

with strict inequalities for at least one pair of combinations of values $x = cy$ and $x' = \bar{c}y$ for the set X of parents of B other than A . It is now readily seen that once a value for the provoking node C is observed, the non-monotonic influence of A on B reduces to a monotonic influence. We say that the observation *resolves* the non-monotonicity of A 's influence on B .

When an observation for its provoking node reduces a non-monotonic influence between two nodes to a monotonic influence, the sign of the resulting influence can be determined from the additive synergy defined for the nodes concerned.

Proposition 4.4 *Let $Q = (G, \Delta)$ be a qualitative probabilistic network. Let A, B, C be nodes in G with $A \rightarrow B, C \rightarrow B \in A(G)$. Then, for any $c_i \in \{c, \bar{c}\}$,*

$$S^{\sim c}(A, B) \wedge Y^\delta(\{A, C\}, B) \wedge C = c_i \implies S^{\delta \otimes \text{sign}[c_i]}(A, B),$$

for all $\delta \in \{+, -, 0, ?\}$, where $\text{sign}[c_i] = +$ if $c_i = c$ and $\text{sign}[c_i] = -$ if $c_i = \bar{c}$.

Proof: Let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Suppose that nodes A and C exhibit a positive additive synergy $Y^+(\{A, C\}, B)$ on node B , that is, we have

$$\Pr(b \mid acy) + \Pr(b \mid \bar{a}\bar{c}y) \geq \Pr(b \mid a\bar{c}y) + \Pr(b \mid \bar{a}cy),$$

for any combination of values y for the set of parents of node B other than A and C . From the non-monotonicity of the influence of A on B and C being its provoker, we now conclude that

$$\Pr(b \mid acy) \geq \Pr(b \mid \bar{a}cy) \quad \text{and} \quad \Pr(b \mid a\bar{c}y) \leq \Pr(b \mid \bar{a}\bar{c}y)$$

must hold for all combinations of values y , with strict inequalities for at least one pair of combinations of values $x = cy$ and $x' = \bar{c}y$ for the set X of parents of B other than A .

Now, suppose that the value c is observed for the provoking node C . We then find

$$\Pr(b \mid ax) \geq \Pr(b \mid \bar{a}x)$$

for any combination of values x , including the observation c , for the set X . We observe that, after resolving the non-monotonicity involved, node A exerts a positive qualitative influence on B . Alternatively, upon observation of \bar{c} , node A exerts a negative influence on B . A negative additive synergy exhibited by A and C on B leads to an analogous result. We conclude that the sign of the qualitative influence from A on B after resolving its non-monotonicity equals the sign-product of the sign of the additive synergy involved and the sign of the observation for the provoking node. \square

The resolution of the non-monotonicity of an influence that is provoked by a single node, is illustrated with an example.

Example 4.5 We consider once again the *Cervical Metastases* network from Figure 4.2. From the probabilities specified for node C it is readily seen that, given a tumour in the upper one-third of a patient's oesophagus, that is, given \bar{l} , node M exerts a negative qualitative influence on C ; given l , M exerts a positive influence on C . In the network, the sign of the influence of M on C after resolution by \bar{l} is computed to be the sign-product of the sign '+' of the additive synergy of M and L on C and the sign '-' of the observation for L , that is, $+ \otimes - = -$. After resolution by l , the sign of the influence of M on C is computed to be $+ \otimes + = +$. \square

For probabilistic inference in a qualitative probabilistic network that explicitly distinguishes between non-monotonic and unknown influences, basically the same algorithm can be used as for probabilistic inference within a regular qualitative network. The only difference lies in the traversal of a non-monotonic qualitative influence: before propagating over a non-monotonic influence, it is investigated whether or not the influence's non-monotonicity is resolved by the available observations. If the non-monotonicity is resolved, the sign of the resulting influence as described above is used in the propagation; otherwise, the ambiguous sign '?' is propagated. The extended algorithm, generalised to an arbitrary set of provokers, is given in Figure 4.3.

Two provokers

So far we have focused our discussion on non-monotonicities provoked by a single node. The concept of provoker, however, can easily be extended to sets of nodes. We first consider a set of two provokers.

Definition 4.6 Let $G = (V(G), A(G))$ be an acyclic digraph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let A, B, C, D be nodes in G with $A \rightarrow B, C \rightarrow B, D \rightarrow B \in A(G)$ such that node A exerts a non-monotonic influence on B . Let $Y = \pi_G(B) \setminus \{A, C, D\}$. Then, the nodes C and D are the provokers of the non-monotonicity of the influence of node A on node B in G , iff there exists a $d_k \in \{d, \bar{d}\}$ such that for some $c_i, c_j \in \{c, \bar{c}\}$, $c_i \neq c_j$,

$$\Pr(b \mid ac_i d_k y) \geq \Pr(b \mid \bar{a}c_i d_k y) \quad \text{and} \quad \Pr(b \mid ac_j d_k y) \leq \Pr(b \mid \bar{a}c_j d_k y),$$

and there exists a $c_k \in \{c, \bar{c}\}$ such that for some $d_i, d_j \in \{d, \bar{d}\}$, $d_i \neq d_j$,

$$\Pr(b \mid ac_k d_i y) \geq \Pr(b \mid \bar{a} c_k d_i y) \quad \text{and} \quad \Pr(b \mid ac_k d_j y) \leq \Pr(b \mid \bar{a} c_k d_j y),$$

for any combination of values y for the set Y .

From this definition we have that for each combination of values z for the set $Z = \{C, D\}$, either $\Pr(b \mid azy) \geq \Pr(b \mid \bar{a}zy)$ holds for any combination of values y for Y , or $\Pr(b \mid azy) \leq \Pr(b \mid \bar{a}zy)$ for any such y . In addition, we note that neither node C nor node D is a provoker of the non-monotonicity by itself. We therefore may need a value for *both* provokers to unambiguously determine the sign of the influence. To illustrate the difference between a non-monotonicity provoked by a single node and a non-monotonicity provoked by two nodes, we consider a node B with three parents, A , C and D . Suppose that for these nodes the following inequalities hold:

$$\begin{aligned} \Pr(b \mid acd) &> \Pr(b \mid \bar{a}cd) & \text{and} & \quad \Pr(b \mid ac\bar{d}) > \Pr(b \mid \bar{a}c\bar{d}) & \text{and} \\ \Pr(b \mid a\bar{c}d) &< \Pr(b \mid \bar{a}\bar{c}d) & \text{and} & \quad \Pr(b \mid a\bar{c}\bar{d}) < \Pr(b \mid \bar{a}\bar{c}\bar{d}), \end{aligned}$$

then we have that

$$\Pr(b \mid acd_i) > \Pr(b \mid \bar{a}cd_i) \quad \text{and} \quad \Pr(b \mid a\bar{c}d_i) < \Pr(b \mid \bar{a}\bar{c}d_i),$$

for any value $d_i \in \{d, \bar{d}\}$ of D . The set of nodes for which a value is needed to unambiguously determine the sign of the influence of node A on node B consists therefore only of node C : node C is the single provoker of the non-monotonicity between A and B . If, on the other hand, the following inequalities hold:

$$\begin{aligned} \Pr(b \mid acd) &> \Pr(b \mid \bar{a}cd) & \text{and} & \quad \Pr(b \mid ac\bar{d}) < \Pr(b \mid \bar{a}c\bar{d}) & \text{and} \\ \Pr(b \mid a\bar{c}d) &< \Pr(b \mid \bar{a}\bar{c}d) & \text{and} & \quad \Pr(b \mid a\bar{c}\bar{d}) > \Pr(b \mid \bar{a}\bar{c}\bar{d}), \end{aligned}$$

then the sign of the difference

$$\Pr(b \mid ac_i d_j) - \Pr(b \mid \bar{a}c_i d_j)$$

depends on *both* the value $c_i \in \{c, \bar{c}\}$ of C and the value $d_j \in \{d, \bar{d}\}$ of D . The set of nodes for which a value is required to unambiguously determine the sign of the influence of node A on node B consists therefore of both nodes C and D . It is readily seen that once values for the provokers C and D are observed, the non-monotonic influence of A on B reduces to a monotonic influence.

Additive synergies can again serve to determine the sign of a resolved influence, once values for the provokers of its non-monotonicity are known. However, in contrast to non-monotonic influences with a single provoker, the additive synergies involved do not always suffice. The following proposition reveals the conditions under which additive synergies do suffice to unambiguously determine the sign of a resolved influence.

Proposition 4.7 *Let $Q = (G, \Delta)$ be a qualitative probabilistic network. Let A, B, C, D be nodes in G with $A \rightarrow B, C \rightarrow B, D \rightarrow B \in A(G)$. Then, for any $c_i \in \{c, \bar{c}\}$, $d_j \in \{d, \bar{d}\}$,*

$$\begin{aligned} S^{\sim\{C,D\}}(A, B) \wedge Y^{\delta_i}(\{A, C\}, B) \wedge Y^{\delta_j}(\{A, D\}, B) \wedge C = c_i \wedge D = d_j \\ \implies S^{(\delta_i \otimes \text{sign}[c_i]) \oplus (\delta_j \otimes \text{sign}[d_j])}(A, B), \end{aligned}$$

for all $\delta_i, \delta_j \in \{+, -, 0, ?\}$, where $\text{sign}[z] = +$ and $\text{sign}[\bar{z}] = -$ for $z \in \{c, d\}$.

Proof: Let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr , and let $Y = \pi_G(B) \setminus \{A, C, D\}$. Suppose that the nodes A and C , respectively A and D , exhibit positive additive synergies $Y^+(\{A, C\}, B)$ and $Y^+(\{A, D\}, B)$ on B , that is, we have

$$\Pr(b \mid acd_i y) + \Pr(b \mid \bar{a}\bar{c}d_i y) \geq \Pr(b \mid a\bar{c}d_i y) + \Pr(b \mid \bar{a}cd_i y),$$

for all values $d_i \in \{d, \bar{d}\}$ of D , and

$$\Pr(b \mid adc_i y) + \Pr(b \mid \bar{a}\bar{d}c_i y) \geq \Pr(b \mid a\bar{d}c_i y) + \Pr(b \mid \bar{a}dc_i y),$$

for all values $c_i \in \{c, \bar{c}\}$ of C ; these inequalities hold for any combination of values y for Y .

From $Z = \{C, D\}$ being the set of provokers for the non-monotonicity of the influence of A on B , we further have that there exist combinations of values z and z' for Z such that

$$\Pr(b \mid azy) \geq \Pr(b \mid \bar{a}zy) \quad \text{and} \quad \Pr(b \mid az'y) \leq \Pr(b \mid \bar{a}z'y), \quad (4.1)$$

for all combinations of values y for Y , with strict inequalities for a least one pair of combinations of values $x = zy$ and $x' = z'y$ for the set $X = \pi_G(B) \setminus \{A\}$. Now, there are four possible combinations of values for both z and z' . For each y there are $2^4 - 2 = 14$ possible sets of four inequalities each that satisfy Equation (4.1). Only ten of these sets of inequalities comply with the fact that neither C nor D is a provoker by itself. Given that both additive synergies are positive, only two sets of inequalities can hold for each combination of values y for Y :

$$\left\{ \begin{array}{l} \Pr(b \mid acdy) \geq \Pr(b \mid \bar{a}cdy), \\ \Pr(b \mid ac\bar{d}y) \leq \Pr(b \mid \bar{a}\bar{c}\bar{d}y), \\ \Pr(b \mid a\bar{c}dy) \leq \Pr(b \mid \bar{a}\bar{c}dy), \\ \Pr(b \mid a\bar{c}\bar{d}y) \leq \Pr(b \mid \bar{a}\bar{c}\bar{d}y) \end{array} \right\} \quad \text{and} \quad \left\{ \begin{array}{l} \Pr(b \mid acdy) \geq \Pr(b \mid \bar{a}cdy), \\ \Pr(b \mid ac\bar{d}y) \geq \Pr(b \mid \bar{a}\bar{c}\bar{d}y), \\ \Pr(b \mid a\bar{c}dy) \geq \Pr(b \mid \bar{a}\bar{c}dy), \\ \Pr(b \mid a\bar{c}\bar{d}y) \leq \Pr(b \mid \bar{a}\bar{c}\bar{d}y) \end{array} \right\}.$$

We observe that for the combinations of values cd and $\bar{c}\bar{d}$ both sets show the same sign for the resolved influence. Given the observations c and d , the sign of the influence of A on B is

$$(\text{sign}[c] \otimes \delta_i) \oplus (\text{sign}[d] \otimes \delta_j) = (+ \otimes +) \oplus (+ \otimes +) = +.$$

Similarly, given the observations \bar{c} and \bar{d} the sign of the influence of A on B is

$$(\text{sign}[\bar{c}] \otimes \delta_i) \oplus (\text{sign}[\bar{d}] \otimes \delta_j) = (- \otimes +) \oplus (- \otimes +) = -.$$

The two sets, however, reveal opposite signs for the combinations of values $c\bar{d}$ and $\bar{c}d$, indicating that the additive synergies do not resolve the non-monotonicity of the influence of A on B . For these combinations of values, the above calculations will result in a '?': the sign of the resolved influence cannot be determined. We conclude that the sign of the qualitative influence of A on B after resolving of its non-monotonicity equals the sign-sum of the sign-products of the sign of the additive synergy involved and the sign of the observation, for the separate provoking nodes from the set of provokers. \square

From the previous proposition we have that additive synergies serve to unambiguously determine the sign of a resolved non-monotonic influence, as long as the sign-products, calculated for each provoker, of the sign of the additive synergy and the sign of the observation, are not conflicting.

Any number of provokers

We have shown that generalising the concept of a single provoker to a set of two provokers is straightforward; we will now further extend the concept to provide for any number of provokers.

Definition 4.8 *Let $G = (V(G), A(G))$ be an acyclic digraph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let $A, B \in V(G)$, $P \subset V(G)$ be nodes in G with $A \rightarrow B$, $P_i \rightarrow B \in A(G)$, for all $P_i \in P$ such that A exerts a non-monotonic influence on B . Let $Y = \pi_G(B) \setminus (\{A\} \cup P)$. Then, P is the provoking set of the non-monotonicity of the influence of node A on node B in G , iff for each $P_i \in P$ there exists a combination of values z for the set of nodes $Z = P \setminus \{P_i\}$ such that for some $p_{ij}, p_{ik} \in \{p_i, \bar{p}_i\}$, $p_{ij} \neq p_{ik}$, we have*

$$\Pr(b \mid ap_{ij}zy) \geq \Pr(b \mid \bar{a}p_{ij}zy) \quad \text{and} \quad \Pr(b \mid ap_{ik}zy) \leq \Pr(b \mid \bar{a}p_{ik}zy),$$

for any combination of values y for Y .

Note that the definitions of a single provoker and of a set of two provokers are special cases of this definition. From the definition we have that no subset of P is a provoking set of the non-monotonicity and that any observed combination of values for P serves to resolve the non-monotonicity.

Without proof, we generalise the proposition that states that the sign of a resolved influence can, under certain conditions, be unambiguously determined from the signs of the resolving observations and the signs of the additive synergies for the nodes involved.

Proposition 4.9 *Let $Q = (G, \Delta)$ be a qualitative probabilistic network, and let A, B , and P be as in the previous definition. Then, for any value $p_{ij} \in \{p_i, \bar{p}_i\}$ of node $P_i \in P$,*

$$S^{\sim P}(A, B) \wedge \bigwedge_{P_i \in P} Y^{\delta_i}(\{A, P_i\}, B) \wedge P = \bigwedge_{P_i \in P} p_{ij} \implies S^{\oplus_{P_i \in P} (\delta_i \otimes \text{sign}[p_{ij}])}(A, B),$$

for all $\delta_i \in \{+, -, 0, ?\}$, where $\text{sign}[p_{ij}] = +$ if $p_{ij} = p_i$, and $\text{sign}[p_{ij}] = -$ if $p_{ij} = \bar{p}_i$.

4.1.2 Probabilistic inference revisited

In the previous section, we briefly described how to extend the basic sign-propagation algorithm from Chapter 3 to allow for propagating observations in a qualitative probabilistic network that explicitly distinguishes between non-monotonic and unknown influences: before propagating a sign over a non-monotonic influence, it is investigated whether or not the influence's non-monotonicity is resolved by the available observations. We recall that a non-monotonicity is ensured to be resolved if an observation is available for each of its provokers. If the non-monotonicity is resolved, the sign of the resolved influence is determined from the signs of the appropriate observations and additive synergies; the resulting sign is subsequently propagated. If observations are not available for each provoker, the ambiguous sign '?' is propagated. The pseudocode summarising this algorithm is given in Figure 4.3.

In our extended algorithm, we attempt to resolve a non-monotonic influence the moment it is encountered during inference. If not enough observations are available to ensure that the influence's non-monotonicity is resolved, a '?' is propagated. Another option would be to propagate

```

procedure PropagateSign(trail, from, to, messagesign):

  sign[to]  $\leftarrow$  sign[to]  $\oplus$  messagesign;
  trail  $\leftarrow$  trail  $\cup$  {to};
  for each active neighbour  $V_i$  of to
  do linksign  $\leftarrow$  sign of (induced) influence between to and  $V_i$ ;
     if linksign  $\equiv$  ' $\sim_P$ '
     then if  $P \subseteq \text{Observed}$ 
        then linksign  $\leftarrow$   $\oplus_j(\text{sign}[P_j] \otimes \delta_j)$ 
           where  $\delta_j$  is determined from  $Y^{\delta_j}(\{to, P_j\}, V_i)$ 
        else linksign  $\leftarrow$  '?';
     messagesign  $\leftarrow$  sign[to]  $\otimes$  linksign;
     if  $V_i \neq \text{trail}$  and sign[ $V_i$ ]  $\neq$  sign[ $V_i$ ]  $\oplus$  messagesign
     then PropagateSign(trail, to, V_i, messagesign)

```

Figure 4.3: The sign-propagation algorithm extended for resolving non-monotonicity of influences.

the ' \sim_P '-sign; it is possible to construct tables for the \oplus -operator and the \otimes -operator that include this sign. The outcome of sign-propagation then includes the ' \sim_P '-sign. If the user does not know the actual values of the provokers, however, the fact that the result is non-monotonic may not be too informative, although the fact that it is non-monotonic *in P* may be insightful. If the user does happen to have a clue about the actual values of the provokers, sign-propagation can be repeated with these values entered as observations, thereby resolving the non-monotonicity. We will not discuss the propagation of ' \sim_P '-signs in further detail.

To conclude, we would like to note that, in our extended algorithm, the sign of a qualitative influence whose non-monotonicity is resolved is determined during inference; the resulting sign is passed on by the algorithm's variable *linksign*. As a result, in propagating subsequent observations the signs of previously resolved influences will be determined over and over again. To prevent these unnecessary calculations, the sign of a non-monotonic influence after resolving its non-monotonicity can also be inserted in the network.

4.1.3 Discussion

A qualitative probabilistic network in essence serves to capture monotonic probabilistic influences between its nodes only. We have argued that it is worthwhile to explicitly capture information about non-monotonic influences, as this information can be exploited in probabilistic inference to prevent, at least to some extent, unnecessarily weak, ambiguous results. To explicitly capture non-monotonicity, a domain expert must identify the influences that are non-monotonic and indicate the provokers of the non-monotonicity. Any combination of observations for the provoking set then serves to resolve the non-monotonicity of the influence. We have proposed the use of additive synergies for subsequently determining the sign of the influence after its resolution. An advantage of using additive synergies is that a domain expert merely has to specify the provokers of a non-monotonicity. The signs of the influence, after resolving its non-monotonicity,

for different combinations of values for the provoking set are not required from the expert: these signs are determined from the signs of the appropriate synergies. A drawback of using additive synergies is that, when the non-monotonicity of an influence is provoked by two or more nodes, the synergies do not always provide enough information to calculate an unambiguous sign for the resolved influence. In Section 4.3 we will show how to circumvent this problem by having the domain expert specify additional information in the form of signs of influences per context.

In this section we focused attention on non-monotonic influences between binary nodes. We will now briefly discuss a generalisation to non-monotonic synergies, non-monotonic intercausal influences, and to non-monotonic influences between non-binary nodes. The definition of non-monotonicity of additive synergies is quite straightforward. Consider, for example, the network shown in Figure 4.4. Suppose that nodes B and C exhibit an additive synergy on node D that is non-monotonic, caused by the single provoker E . Then, for some $e_i, e_j \in \{e, \bar{e}\}$, $e_i \neq e_j$, this non-monotonicity is described by

$$\begin{aligned} \Pr(d \mid bce_i) + \Pr(d \mid \bar{b}\bar{c}e_i) &> \Pr(d \mid b\bar{c}e_i) + \Pr(d \mid \bar{b}ce_i) \quad \text{and} \\ \Pr(d \mid bce_j) + \Pr(d \mid \bar{b}\bar{c}e_j) &< \Pr(d \mid b\bar{c}e_j) + \Pr(d \mid \bar{b}ce_j). \end{aligned}$$

At present, we see no way of resolving the non-monotonicity of an additive synergy. Also, as additive synergies are not used upon inference for any other purpose than for resolving non-monotonic influences, we refrain from further investigation.

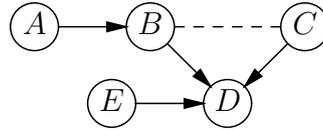


Figure 4.4: An example network.

Non-monotonicity of product synergies and intercausal influences is more interesting, yet also more complicated. We again consider Figure 4.4, showing a product synergy for the nodes B and C with regard to values for their common child D . As the intercausal influence induced by a product synergy is quite similar to a regular qualitative influence, non-monotonicity of the intercausal influence between two nodes B and C would, on first thought, be provoked by another parent of either B or C , such as node A in Figure 4.4. However, the intercausal influence between the nodes B and C is induced by the product synergy exhibited by these nodes with respect to a specific value of node D , and D is independent of A given B . The product synergy can therefore not be non-monotonic in A and hence cannot induce an intercausal influence that is non-monotonic in A . Another possibility is that a non-monotonic intercausal influence is described by a non-monotonic product synergy whose non-monotonicity is provoked by another parent of the node on which the synergy is exhibited, such as node E in the figure. The non-monotonicity of the product synergy can then be defined in a similar way as for additive synergies. Note that only product synergies of type II and their induced intercausal influences can be non-monotonic. We suspect that additive synergies can again be used to resolve the non-monotonicity of an intercausal influence, but further research is still required.

So far, we have focused our discussion on binary nodes. We recall from the discussion in Chapter 3 that our assumption of all nodes being binary is not a restrictive one, since the basic

sign-propagation algorithm does not allow for a finer level of detail. We would, however, like to give an indication of how to extend the concept of non-monotonicity and its resolution to non-binary nodes. We have from our definition of provoker for binary nodes that a non-monotonic influence becomes monotonic once values for its provokers are available. Building upon this idea, we should have for a non-monotonic influence of a non-binary node B on a non-binary node D that is provoked by binary node C , as in Figure 4.4, that for all values d_i of D , and all values $b_j > b_k$ of B , either

$$\Pr(D \geq d_i \mid b_j cx) \geq \Pr(D \geq d_i \mid b_k cx) \text{ and } \Pr(D \geq d_i \mid b_j \bar{c}x) \leq \Pr(D \geq d_i \mid b_k \bar{c}x), \text{ or}$$

$$\Pr(D \geq d_i \mid b_j cx) \leq \Pr(D \geq d_i \mid b_k cx) \text{ and } \Pr(D \geq d_i \mid b_j \bar{c}x) \geq \Pr(D \geq d_i \mid b_k \bar{c}x),$$

with strict inequalities for at least one pair of combinations of values $y = cx$ and $y' = \bar{c}x$ for $Y = \{X\} \cup \{C\}$ where $X = E$. Note that we still assume the provoking node to be binary. The method of resolving non-monotonic influences through the use of additive synergies now applies straightforwardly because there are only two possible values for the provoking node; we therefore know that the resolved influence is negative for one value and positive for the other. With a set of provokers, we have one additive synergy per provoker, which again provides us with only two possibilities. A possible idea for handling non-binary provokers is to split each provoker into a set of binary nodes, but this idea awaits further investigation.

4.2 Enhanced qualitative probabilistic networks

Ambiguous signs during inference with a qualitative probabilistic network not only result from the presence of non-monotonic influences, but are also generated from trade-offs in the network. A network models a trade-off if it contains multiple parallel trails such that the signs of the influences along these trails are conflicting. To adequately deal with trade-offs, we have designed the formalism of enhanced qualitative probabilistic networks in which we distinguish between strong and weak qualitative influences. By introducing a qualitative notion of relative strength, several trade-offs can be resolved during inference by building upon the idea that strong influences dominate over conflicting weak influences. For this purpose, we have generalised the sign-propagation algorithm for regular qualitative networks to apply to enhanced networks.

4.2.1 The enhanced formalism

In a *quantitative* probabilistic network, the net result from conflicting influences between a node A and a node B along multiple parallel trails is computed from the conditional probabilities specified for the nodes on the trails. The conditional probabilities can be regarded as specifying the strengths of the influences in a quantitative network. The coarse level of representation detail of a qualitative probabilistic network, however, does not provide for an indication of the strength of an influence. As a consequence, if a qualitative network models a trade-off, the composition of the conflicting influences will result in the generation of a ‘?’ during inference; we say that the trade-off remains *unresolved*. The following example illustrates a trade-off.

Example 4.10 We consider the probabilistic *Radiotherapy* network and its qualitative abstraction in Figure 4.5. The network represents a highly simplified fragment of the prognostic part of the oesophagus network. Node L models the life-expectancy of a patient after therapy, where l indicates that the patient will survive for at least 6 weeks. Node T models the therapy instilled;

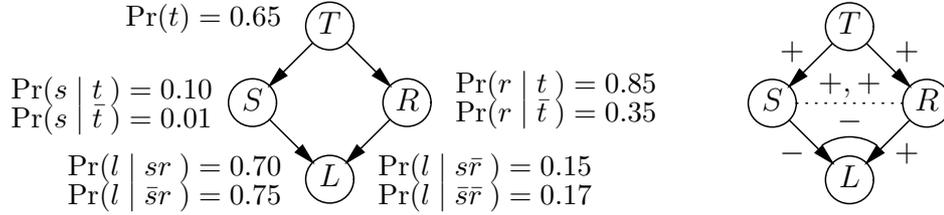


Figure 4.5: The *Radiotherapy* network.

we focus on whether or not a patient receives radiotherapy, modelled by t and \bar{t} , respectively. The effect to be attained from radiotherapy is the reduction r of the tumour in the patient's oesophagus, modelled by node R . A complication associated with radiotherapy is a build up of scar tissue, called stenosis, causing narrowing of the oesophagus; the presence or absence of stenosis is modelled by node S . If the patient receives radiotherapy, his life expectancy may decrease due to stenosis. His life expectancy will on the other hand increase if the tumour is reduced. From the probabilities in the quantitative network, we have that the effect of stenosis on life expectancy is much smaller than the effect of tumour reduction. The fact that the influence of S on L is much weaker than the influence of R on L is, however, not apparent from the qualitative abstraction. During inference, therefore, the trade-off cannot be resolved. \square

To allow for resolving trade-offs in a qualitative way, we enhance the formalism of qualitative probabilistic networks by introducing a notion of relative strength of influences. If, for example, a positive influence is known to be stronger than a conflicting negative one, we may then conclude that the combined influence is positive, thereby resolving the trade-off. In an *enhanced qualitative probabilistic network*, we distinguish to this end between strong and weak influences. Intuitively, a strong influence of a node A on a node B is an influence that is stronger than any weak influence in the network, that is

$$|\Pr(b | ax) - \Pr(b | \bar{a}x)| \geq |\Pr(d | cy) - \Pr(d | \bar{c}y)|,$$

for all nodes C and D with a weak influence between them; the inequalities should hold for each combination of values x for the set of parents of B other than A and for each combination of values y for the set of parents of D other than C . The basic idea now is to partition the set of all influences of a network into two disjoint sets of influences in such a way that any influence from the one subset is stronger than any influence from the other subset. To this end, a *cut-off value* α is introduced. This value serves to partition the set of all qualitative influences into a set of influences that capture an absolute difference in probabilities larger than α and a set of influences that model an absolute difference smaller than α . An influence from the former subset will be termed a *strong influence*; an influence from the latter subset will be termed a *weak influence*.

Definition 4.11 Let $G = (V(G), A(G))$ be an acyclic digraph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let A, B be nodes in G , with $A \rightarrow B \in A(G)$. Let $X = \pi(B) \setminus \{A\}$ and let α be a cut-off value. The influence of node A on node B is strongly positive, denoted $S^{++}(A, B)$, iff it is a positive qualitative influence with

$$\Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \geq \alpha,$$

for any combination of values x for the set X . The influence of node A on node B is weakly positive, denoted $S^+(A, B)$, iff it is a positive qualitative influence with

$$\Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \leq \alpha,$$

for any combination of values x for X .

Strongly negative qualitative influences, denoted S^{--} , and weakly negative qualitative influences, denoted S^- , are defined analogously; zero qualitative influences and ambiguous qualitative influences are defined as in regular qualitative probabilistic networks.

We would like to note that, in our enhanced formalism, the meaning of the sign of a qualitative influence has slightly changed. While in a regular qualitative probabilistic network, the sign of an influence represents the sign of a difference in probabilities only, in an enhanced qualitative network a sign in addition captures the relative magnitude of the difference.

Upon abstracting a quantified probabilistic network to an enhanced qualitative probabilistic network, the cut-off value α needs to be chosen explicitly. This cut-off value will typically vary from application to application. We would like to note that it is always possible to choose a cut-off value, as the value $\alpha = 1$ yields a trivial partitioning of the set of influences. In real-life applications of enhanced qualitative probabilistic networks, however, a cut-off value does *not* have to be established explicitly. The partitioning into strong and weak influences then is elicited directly from the domain experts involved in the construction of the network.

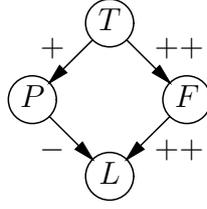
Example 4.12 We consider again the *Radiotherapy* network from Figure 4.5. Suppose that we choose for our cut-off value $\alpha = 0.45$. For the influence of node T on node R , we now find that

$$\Pr(r \mid t) - \Pr(r \mid \bar{t}) \geq 0, \quad \text{and} \quad |\Pr(r \mid t) - \Pr(r \mid \bar{t})| = 0.50 \geq \alpha.$$

We therefore conclude that $S^{++}(T, R)$. We further find that $S^+(T, S)$, $S^-(S, L)$, and $S^{++}(R, L)$. The resulting enhanced qualitative probabilistic network, showing only the qualitative influences involved, is depicted in Figure 4.6. \square

4.2.2 Properties of enhanced qualitative influences

We recall from Section 3.2 that the sign-propagation algorithm for probabilistic inference builds on the idea of propagating signs throughout a qualitative network and combining them with the \otimes - and \oplus -operators. We further recall that the algorithm exploits the properties of symmetry, transitivity, and composition of influences. To generalise the idea of sign-propagation to inference with an enhanced qualitative probabilistic network, we begin by enhancing the \otimes - and \oplus -operators to provide for the properties of transitivity and composition of strong and weak influences; after doing so, we focus on the property of symmetry.

Figure 4.6: The enhanced *Radiotherapy* network.

Enhancing the \otimes -operator

For propagating qualitative signs along trails of nodes in an enhanced qualitative probabilistic network, we enhance the \otimes -operator that is defined for this purpose for regular qualitative networks, to apply to strong and weak influences. In a regular qualitative probabilistic network, an influence basically captures a difference between probabilities. Combining two influences with the property of transitivity amounts to determining the sign of the product of two such differences. In our formalism of enhanced qualitative probabilistic networks, we have associated an explicit notion of relative strength with influences. It will be evident that these relative strengths need to be taken into consideration when multiplying signs.

To address the sign-product of two signs in an enhanced qualitative probabilistic network, we consider the network fragment shown in Figure 4.7. The fragment includes the nodes A , B and

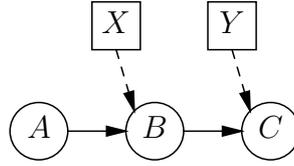


Figure 4.7: A fragment of a network.

C ; in addition, X denotes the set of all parents of B other than A , and Y is the set of all parents of C other than B . From the proof of Proposition 3.9 in Section 3.1, we have that

$$\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) = (\Pr(c \mid by) - \Pr(c \mid \bar{b}y)) \cdot (\Pr(b \mid ax) - \Pr(b \mid \bar{a}x)), \quad (4.2)$$

for all combinations of values x for the set of nodes X and y for the set Y . The differences $\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy)$ give an indication of the relative strength of the indirect influence of A on C . We now consider the possible combinations of signs for the influences associated with the arcs in the network fragment under consideration, and their sign-products. Suppose that both qualitative influences are strongly positive, that is, we have $S^{++}(A, B)$ and $S^{++}(B, C)$. Let α be the cut-off value used for distinguishing between strong and weak influences. From Equation (4.2) stated above, we find that

$$\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) \geq \alpha^2$$

for any combination of values xy for the set of nodes $X \cup Y$. Since $\alpha \leq 1$, we have $\alpha^2 \leq \alpha$. Upon multiplying the signs of two strong influences, therefore, a sign results that expresses an indirect influence that is not necessarily stronger than a direct weakly positive influence. Similar

observations apply for strongly negative influences. Now suppose that both qualitative influences in the network fragment from Figure 4.7 are weakly positive, that is, we have $S^+(A, B)$ and $S^+(B, C)$. For the indirect influence of node A on node C , we find that

$$0 \leq \Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) \leq \alpha^2$$

for any combination of values xy for the set $X \cup Y$. Similar observations apply for weakly negative influences.

From the previous observations, we conclude that while the indirect influence resulting from the product of two strong influences cannot be compared to a direct weak influence, the indirect influence is always at least as strong as an indirect influence resulting from the product of two weak influences. To provide for comparing indirect qualitative influences along different trails with respect to their strength, as required for trade-off resolution, we therefore will retain the length of the trail over which influences have been multiplied. To this end, we augment every influence's sign by a superscript, called the sign's *multiplication index*. We would like to note that the multiplication index is only used for the purpose of computation and we do not intend to output signs augmented with these indices to the user. The following definition describes the meaning of an influence with an augmented sign.

Definition 4.13 Let $G = (V(G), A(G))$ be an acyclic digraph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let A and B be nodes in G and let t be a trail from A to B in G . Let $X = (\bigcup_{C \in V(t) \setminus \{A\}} \pi_G(C) \setminus V(t))$. Then, the qualitative influence of node A on node B along trail t is strongly positive with multiplication index i , $i \in \mathbb{N}$, written $\hat{S}^{++^i}(A, B, t)$, iff

$$\Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \geq \alpha^i$$

for every combination of values x for the set of nodes X . The qualitative influence of A on B along trail t is weakly positive with multiplication index i , $i \in \mathbb{N}$, written $\hat{S}^{+^i}(A, B, t)$, iff

$$0 \leq \Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \leq \alpha^i$$

for every combination of values x for the set of nodes X .

Weakly and strongly negative influences with a multiplication index are again defined analogously. The signs of the influences associated with the arcs of the digraph of an enhanced qualitative network are interpreted as having a multiplication index equal to 1. Building on the concept of multiplication index, Table 4.1 now defines the enhanced \otimes -operator. From the table, it is readily seen that the '+', '-', '0', and '?' signs combine as in a regular qualitative probabilistic network; the only difference is the handling of the multiplication indices. The following two lemmas show that the sign-product, as specified in Table 4.1, of two signs δ_i and δ_j indeed corresponds to the sign of the transitive combination of the influences with the signs δ_i and δ_j . The previous observations with respect to strongly positive influences are summarised in the following lemma.

\otimes	$++^j$	$+^j$	0	$-^j$	$---^j$?
$++^i$	$++^{i+j}$	$+^j$	0	$-^j$	$---^{i+j}$?
$+^i$	$+^i$	$+^{i+j}$	0	$-^{i+j}$	$-^i$?
0	0	0	0	0	0	0
$-^i$	$-^i$	$-^{i+j}$	0	$+^{i+j}$	$+^i$?
$---^i$	$---^{i+j}$	$-^j$	0	$+^j$	$++^{i+j}$?
?	?	?	0	?	?	?

Table 4.1: The enhanced \otimes -operator.

Lemma 4.14 *Let $Q = (G, \Delta)$ be an enhanced qualitative probabilistic network. Let A, B , and C be nodes in G , and let t_i and t_j be trails in G from A to B and from B to C , respectively, such that their trail concatenation $t_i \circ t_j$ is sinkless. Then,*

$$\hat{S}^{++^i}(A, B, t_i) \wedge \hat{S}^{++^j}(B, C, t_j) \implies \hat{S}^{++^{i+j}}(A, C, t_i \circ t_j).$$

From this lemma we see that for strongly positive influences the enhanced \otimes -operator indeed correctly captures the sign $++^i \otimes ++^j = ++^{i+j}$ of their transitive combination. Similar observations hold for the transitive combination of two weak or two strong influences, be they positive or negative.

The following lemma provides for the transitive combination of a weak and a strong influence.

Lemma 4.15 *Let Q, A, B, C, t_i and t_j be as in the previous lemma. Then,*

$$\hat{S}^{+^i}(A, B, t_i) \wedge \hat{S}^{++^j}(B, C, t_j) \implies \hat{S}^{+^i}(A, C, t_i \circ t_j).$$

Proof: Let Pr be a joint probability distribution on $V(G)$ such that G is an I-map for Pr . Let α be the cut-off value used for distinguishing between strong and weak influences. The weakly positive influence $\hat{S}^{+^i}(A, B, t_i)$ of node A on node B expresses that

$$0 \leq \text{Pr}(b \mid ax) - \text{Pr}(b \mid \bar{a}x) \leq \alpha^i,$$

for every combination of values x for the set X of relevant ancestors of node B . The strongly positive qualitative influence $\hat{S}^{++^j}(B, C, t_j)$ of node B on node C further expresses that

$$\text{Pr}(c \mid by) - \text{Pr}(c \mid \bar{b}y) \geq \alpha^j,$$

for every combination of values y for the set Y of relevant ancestors of node C . We observe that $\alpha^j \leq 1$. For the influence of A on C along the trail concatenation $t_i \circ t_j$, we now find that

$$\text{Pr}(c \mid axy) - \text{Pr}(c \mid \bar{a}xy) \leq \alpha^i \cdot \alpha^j,$$

for every combination of values xy for the set $X \cup Y$. From $\alpha^i \cdot \alpha^j \leq \alpha^i$, we conclude that $\hat{S}^{+^i}(A, C, t_i \circ t_j)$. \square

From the above lemma we conclude that for a weakly and a strongly positive influence the enhanced \otimes -operator indeed correctly captures the sign $+^i \otimes ++^j = +^i$ of their transitive combination. Similar observations hold for the transitive combination of any weak and any strong influence, be they positive or negative.

Enhancing the \oplus -operator

For combining multiple qualitative influences between two nodes along parallel trails in an enhanced qualitative network, we enhance the \oplus -operator that is defined for this purpose for regular qualitative probabilistic networks, to apply to strong and weak influences. When addressing the enhanced \otimes -operator, we have argued that the product of two influences may yield an indirect influence that is weaker than the influences it is built from. We will now see that the sum of two influences, in contrast, may result in a stronger influence.

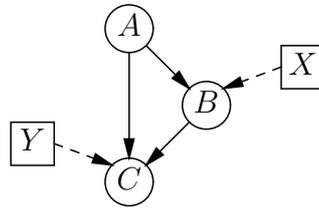


Figure 4.8: Another network fragment.

To address the sign-sum of two signs in an enhanced qualitative probabilistic network, we consider the network fragment shown in Figure 4.8. The fragment includes the nodes A , B and C ; in addition, X denotes the set of all parents of B other than A , and Y is the set of all parents of C other than A and B . From the proof of Proposition 3.10 in Section 3.1, we recall that

$$\begin{aligned} \Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) &= (\Pr(c \mid aby) - \Pr(c \mid \bar{a}\bar{b}y)) \cdot \Pr(b \mid ax) + \Pr(c \mid \bar{a}\bar{b}y) \\ &\quad - (\Pr(c \mid \bar{a}by) - \Pr(c \mid \bar{a}\bar{b}y)) \cdot \Pr(b \mid \bar{a}x) - \Pr(c \mid \bar{a}\bar{b}y), \end{aligned} \quad (4.3)$$

for all combinations of values x for the set of nodes X and y for the set Y . The differences $\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy)$ give an indication of the difference between the relative strengths of the direct influence and the indirect influence between node A and node C . If all arcs in the figure are associated with a weakly positive influence, we find that $\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) \leq \alpha + \alpha^2$; we will prove this property shortly. From the inequality, we observe that the parallel composition of two weakly positive influences may result in an indirect influence that is stronger than a direct weakly positive influence; note that the strength of the net influence is expressed as a sum of different powers of the cut-off value. The addition of a positive and a negative influence will result in an influence that is weaker than the strongest influence added, and stronger than the weakest influence added; the strength of the net influence can then be expressed as a difference in powers of the cut-off value.

As the relative strength of a composite influence depends on the relative strengths of the influences it is built from, the multiplication indices of the signs of the added influences have to be incorporated into the multiplication index of the resulting influence's sign. To this end, the sign of the composite influence is augmented with a *list* of multiplication indices. Since the

power of the cut-off value α can only be positive, we can indicate subtraction of powers of α by a negative multiplication index. For example, the sign of an influence with strength less than or equal to $\alpha - \alpha^2$ will be denoted by $+^{1,-2}$. We now formally define the meaning of a list of multiplication indices.

Definition 4.16 Let $G = (V(G), A(G))$ be an acyclic digraph and let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let A and B be nodes in G and let $t = t_1 \parallel \dots \parallel t_k$ be a composite trail from A to B in G . Let $X = (\bigcup_{C \in V(t) \setminus \{A\}} \pi_G(C) \setminus V(t))$. Then, the qualitative influence of node A on node B along trail t is strongly positive with the list of multiplication indices i_1, \dots, i_k , $k \geq 1$, $i_j \in \mathbb{Z}$, $j = 1, \dots, k$, written $\hat{S}^{++^{i_1, \dots, i_k}}(A, B, t)$, iff

$$\Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \geq \sum_{i_j > 0} \alpha^{i_j} - \sum_{i_j < 0} \alpha^{|i_j|}$$

for any combination of values x for the set of nodes X . The qualitative influence of node A on node B along trail t is weakly positive with the list of multiplication indices i_1, \dots, i_k , written $\hat{S}^{+^{i_1, \dots, i_k}}(A, B, t)$, iff

$$0 \leq \Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \leq \sum_{i_j > 0} \alpha^{i_j} - \sum_{i_j < 0} \alpha^{|i_j|}$$

for any combination of values x for X .

Strongly and weakly negative influences with a list of multiplication indices are again defined analogously.

\oplus	$++^j$	$+^j$	0	$-^j$	$--^j$?	
$++^i$	$++^{i,j}$	$++^i$	$++^i$	a)	?	?	a) $++^{i,-j}$, if $i \leq j$; ?, otherwise
$+^i$	$++^j$	$+^{i,j}$	$+^i$?	d)	?	b) $++^{-i,j}$, if $j \leq i$; ?, otherwise
0	$++^j$	$+^j$	0	$-^j$	$--^j$?	c) $--^{i,-j}$, if $i \leq j$; ?, otherwise
$-^i$	b)	?	$-^i$	$-^{i,j}$	$--^j$?	d) $--^{-i,j}$, if $j \leq i$; ?, otherwise
$--^i$?	c)	$--^i$	$--^i$	$--^{i,j}$?	
?	?	?	?	?	?	?	

Table 4.2: The enhanced \oplus -operator.

Building upon the concept of a list of multiplication indices, Table 4.2 defines the enhanced \oplus -operator. From the table, it is readily seen that the ‘+’, ‘-’, ‘0’, and ‘?’ signs combine as in a regular qualitative probabilistic network; the only difference is in the handling of the multiplication indices. The following three lemmas show that the sign-sum, as specified in Table 4.2, of two signs δ_i and δ_j indeed corresponds to the sign of the composition of the influences with the signs δ_i and δ_j .

Lemma 4.17 Let $Q = (G, \Delta)$ be an enhanced qualitative probabilistic network. Let A, C be nodes in G and let t_i and t_j be parallel trails in G from A to C . Then

$$\hat{S}^{+^i}(A, C, t_i) \wedge \hat{S}^{+^j}(A, C, t_j) \implies \hat{S}^{+^{i,j}}(A, C, t_i \parallel t_j).$$

Proof: Let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Let α be the cut-off value used for distinguishing between strong and weak influences. For ease of exposition, we assume that trail t_i consists of a single arc and trail t_j consists of the arcs $A \rightarrow B, B \rightarrow C$ for some node B ; we recall that with each arc is associated an influence with multiplication index 1. Equation (4.3) gives the net influence of node A on node C ; we can write this equation as the difference between two functions f and g :

$$\begin{aligned} \Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) &= (\Pr(c \mid aby) - \Pr(c \mid \bar{a}by)) \cdot \Pr(b \mid ax) + \Pr(c \mid \bar{a}by) \\ &\quad - (\Pr(c \mid \bar{a}by) - \Pr(c \mid \bar{a}\bar{b}y)) \cdot \Pr(b \mid \bar{a}x) - \Pr(c \mid \bar{a}\bar{b}y) \\ &= f(\Pr(b \mid ax)) - g(\Pr(b \mid \bar{a}x)), \end{aligned}$$

for all value combinations x and y for the set X of parents of B other than A and the set Y of parents of C other than A and B , respectively. If we assume that the positive influence along trail t_j is composed of two positive influences, then we have that the functions f and g are linearly increasing functions, similar to the ones shown in Figure 4.9; the fact that in the figure the gradient of function f is larger than that of function g is an arbitrary choice. From the positive direct influence of node A on node C we have that $f(0) \geq g(0)$ and $f(1) \geq g(1)$; the functions, therefore, do not intersect. If the influences along trail t_j are negative, the functions f and g will be decreasing and similar observations will apply.

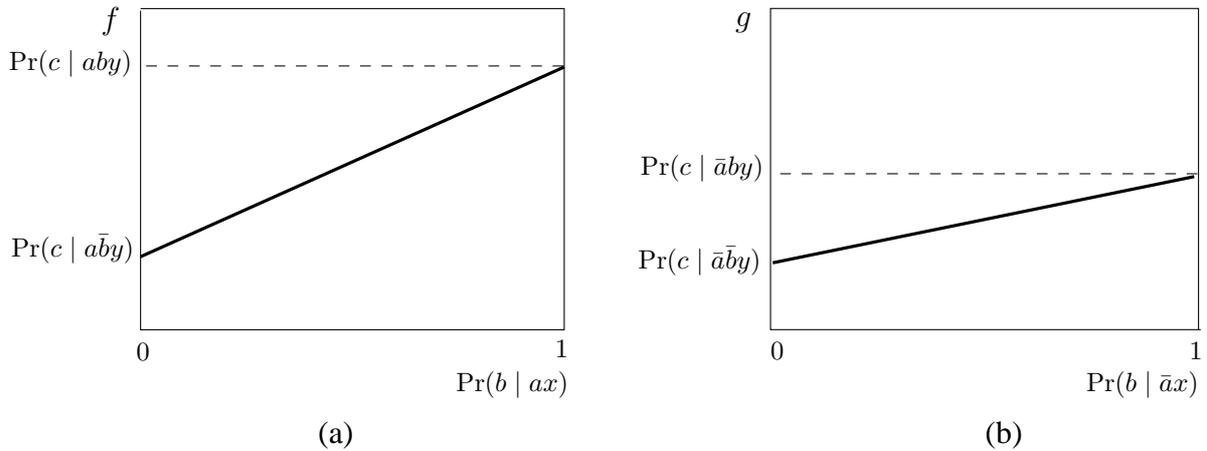


Figure 4.9: Possible functions f , (a), and g , (b).

To determine the sign of the composite influence of node A on node C , we have to consider the sign of the difference between the functions f and g . Although the functions f and g are expressed in terms of different parameters, these parameters cannot be varied independently as their difference is restricted by the sign of the qualitative influence of node A on node B . It is therefore allowed to compare the function values of f and g for different values from their respective domains. For ease of comparison, we depict for this purpose the two functions f and g in a single figure, Figure 4.10.

We now assume that the positive influence along trail t_j is composed of two positive influences, for which we have three possibilities: $S^{+1}(A, B)$ and $S^{++1}(B, C)$, or $S^{++1}(A, B)$ and

$S^{+1}(B, C)$, or $S^{+1}(A, B)$ and $S^{+1}(B, C)$. Here, we only consider the latter case; proofs for the other two cases are similar. As the influence of node A on node B is weakly positive, we have that $0 \leq \Pr(b | ax) - \Pr(b | \bar{a}x) \leq \alpha$. Therefore, when investigating the difference between f and g , we have to satisfy the following constraints:

- the domain value $\Pr(b | ax)$ for f should be greater than or equal to the domain value $\Pr(b | \bar{a}x)$ for g ;
- the difference between the two domain values may not be greater than α .

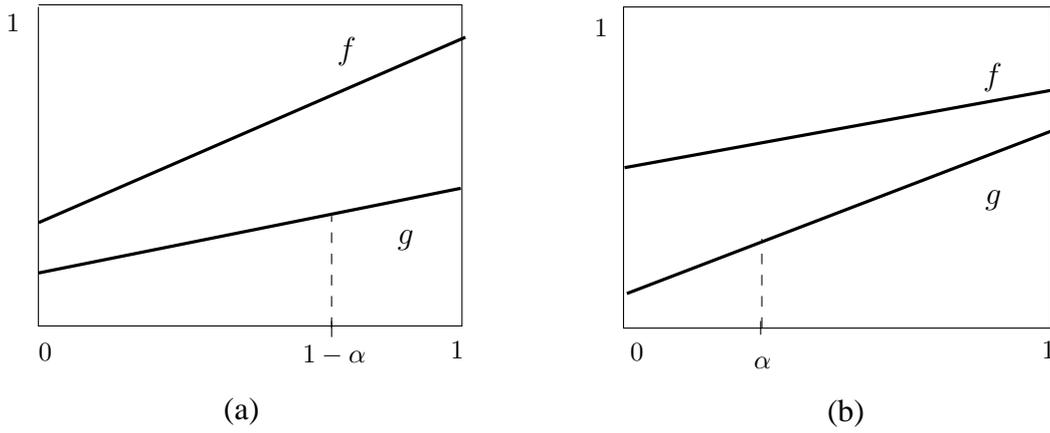


Figure 4.10: The functions $f(\Pr(b | ax))$ and $g(\Pr(b | \bar{a}x))$ depicted in a single graph, with $\text{gradient}(f) > \text{gradient}(g)$, (a), and $\text{gradient}(g) > \text{gradient}(f)$, (b).

We now show that, under these constraints, the difference $f(\Pr(b | ax)) - g(\Pr(b | \bar{a}x))$ is greater than or equal to zero and that the maximal difference has an upper bound expressed in terms of α . To this end, we consider the graph from Figure 4.10(a); similar observations hold for the graph from Figure 4.10(b). Under the constraints mentioned above, we have that the minimal difference between $f(\Pr(b | ax))$ and $g(\Pr(b | \bar{a}x))$ is attained for $f(0)$ and $g(0)$. We find that

$$\Pr(c | axy) - \Pr(c | \bar{a}xy) = f(0) - g(0) = \Pr(c | \bar{a}by) - \Pr(c | \bar{a}\bar{b}y) \geq 0.$$

The minimal difference is positive as a result of the direct influence of A on C being positive. The sign of the net influence of node A on node B is therefore positive. The maximal difference between $f(\Pr(b | ax))$ and $g(\Pr(b | \bar{a}x))$ is attained for $f(1)$ and $g(1 - \alpha)$. Again exploiting the information that the signs of the direct influences are all weakly positive, this maximal difference is:

$$\begin{aligned} \Pr(c | axy) - \Pr(c | \bar{a}xy) &= f(1) - g(1 - \alpha) \\ &= \Pr(c | aby) - \Pr(c | \bar{a}\bar{b}y) - (\Pr(c | \bar{a}by) - \Pr(c | \bar{a}\bar{b}y)) \cdot (1 - \alpha) \\ &= \Pr(c | aby) - \Pr(c | \bar{a}by) + \alpha \cdot (\Pr(c | \bar{a}by) - \Pr(c | \bar{a}\bar{b}y)) \\ &\leq \alpha + \alpha \cdot \alpha = \alpha + \alpha^2. \end{aligned}$$

We conclude that the composite influence of node A on node C is weakly positive with the list of multiplication indices 1, 2, that is, we conclude that $\hat{S}^{+1,2}(A, C, t_i \parallel t_j)$. \square

From the above lemma, we conclude that for two weakly positive influences the enhanced \oplus -operator correctly captures the sign $+^i \oplus +^j = +^{i,j}$ of their composition. Similar observations hold for the composition of two strongly positive signs, two strongly negative signs and two weakly negative signs.

The next lemma addresses the combination of a strongly positive and a weakly positive influence.

Lemma 4.18 *Let Q, A, C and t_i and t_j be as in the previous lemma. Then,*

$$\hat{S}^{++^i}(A, C, t_i) \wedge \hat{S}^{+^j}(A, C, t_j) \implies \hat{S}^{++^i}(A, C, t_i \parallel t_j).$$

Proof: The proof proceeds in a similar fashion as the proof of Lemma 4.17, up to the point where the minimal difference between $f(\Pr(b \mid ax))$ and $g(\Pr(b \mid \bar{a}x))$ is considered. The minimal difference between $f(\Pr(b \mid ax))$ and $g(\Pr(b \mid \bar{a}x))$ is again attained for $f(0)$ and $g(0)$ and we thus have,

$$\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) = f(0) - g(0) = \Pr(c \mid \bar{a}by) - \Pr(c \mid \bar{a}\bar{b}y).$$

Since the direct influence of node A on node C is strongly positive, we have that $\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) \geq \alpha$. We conclude that the composite influence of node A on node C is strongly positive with multiplication index 1, that is, we conclude that $\hat{S}^{++^1}(A, C, t_i \parallel t_j)$. \square

From the above lemma, we deduce that for a weakly and a strongly positive influence the enhanced \oplus -operator correctly captures the sign $+^i \oplus +^j = +^{i,j}$ of their composition. Similar observations hold for the composition of a strongly negative and a weakly negative influence.

The next lemma provides for the combination of conflicting influences using the enhanced \oplus -operator.

Lemma 4.19 *Let Q, A, C, t_i and t_j be as in the previous lemma and let i and j be multiplication indices such that $i \leq j$. Then,*

$$\hat{S}^{++^i}(A, C, t_i) \wedge \hat{S}^{-^j}(A, C, t_j) \implies \hat{S}^{++^{i,-j}}(A, C, t_i \parallel t_j).$$

Proof: Let \Pr, α , trail t_i and trail t_j be as in the proof of Lemma 4.17. Equation (4.3) once again gives the net influence of node A on node C . As in the proof of Lemma 4.17, we construct two functions f and g . Depending on the sign of the influence of node B on node C , we once again have that the functions f and g are either linearly increasing functions as in Figure 4.10 or linearly decreasing functions. From here on we consider the graph from Figure 4.10(a); similar observations apply to the situation in which f and g converge, as illustrated in Figure 4.10(b).

We assume that the negative influence along trail t_j is composed of a negative influence of A on B and a positive influence of B on C , that is, we have either

1. $S^{-1}(A, B)$ and $S^{+1}(B, C)$, or
2. $S^{-1}(A, B)$ and $S^{+1}(B, C)$, or
3. $S^{-1}(A, B)$ and $S^{++^1}(B, C)$.

The indirect transitive influence of node A on node C now has sign ‘ -2 ’ in situation (1) and sign ‘ -1 ’ in the situations (2) and (3). Similar observations hold for a positive influence of node A on node B and a negative influence of B on node C .

We first consider the situations (1) and (3) described above. Since the influence of node A on node B is weakly negative, we have to satisfy $\Pr(b \mid ax) \leq \Pr(b \mid \bar{a}x)$ and $\Pr(b \mid \bar{a}x) - \Pr(b \mid ax) \leq \alpha$ when investigating the difference between f and g . From Equation (4.3), we find that the minimal difference between f and g is now attained for $f(0)$ and $g(\alpha)$; this minimal difference is:

$$\begin{aligned} \Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) &= f(0) - g(\alpha) \\ &= \Pr(c \mid \bar{a}by) - \Pr(c \mid \bar{a}\bar{b}y) - (\Pr(c \mid \bar{a}by) - \Pr(c \mid \bar{a}\bar{b}y)) \cdot \alpha. \end{aligned}$$

In situation (1) we have that $\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) \geq \alpha - \alpha \cdot \alpha$ and we conclude that the net influence of node A on node C is strongly positive with the list of multiplication indices 1, -2 , that is, $\hat{S}^{++^{1,-2}}(A, C, t_i \parallel t_j)$. For situation (3) we find that $\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) \geq \alpha - 1 \cdot \alpha$ and we conclude $\hat{S}^{++^{1,-1}}(A, C, t_i \parallel t_j)$.

We now consider situation (2) described above. Given the strongly negative influence of node A on node B , we must satisfy: $\Pr(b \mid ax) \leq \Pr(b \mid \bar{a}y)$ and $\Pr(b \mid \bar{a}y) - \Pr(b \mid ax) \geq \alpha$ when investigating the difference of the functions f and g . From Equation (4.3), we find that the minimal difference between f and g is now attained for $f(0)$ and $g(1)$; this minimal difference is:

$$\begin{aligned} \Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) &= f(0) - g(1) \\ &= \Pr(c \mid \bar{a}by) - \Pr(c \mid \bar{a}b\bar{y}). \end{aligned}$$

From the strongly positive direct influence of node A on node C , we have that $\Pr(c \mid \bar{a}by) - \Pr(c \mid \bar{a}\bar{b}y) \geq \alpha$ and from the weakly positive influence of node B on node C we find $\Pr(c \mid \bar{a}by) - \Pr(c \mid \bar{a}\bar{b}y) \leq \alpha$. We thus have that $\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) \geq \alpha - \alpha$ and we conclude that the net influence of node A on node C is strongly positive with the list of multiplication indices 1, -1 , that is, $\hat{S}^{++^{1,-1}}(A, C, t_i \parallel t_j)$. \square

From the above lemma, we conclude that for a strongly positive influence with multiplication index i and a weakly negative influence with multiplication index j , $i \leq j$, the enhanced \oplus -operator correctly captures the sign $++^i \oplus -^j = ++^{i,-j}$ of their composition. Similar observations apply to other combinations of strong and weak conflicting influences. From the lemma we have that, under certain conditions, the composition of conflicting strong and weak influences leads to an unambiguous result. The enhanced \oplus -operator thus serves to resolve certain trade-offs.

We would like to note that the enhanced \oplus -operator defined in Table 4.2 is non-associative. We find, for example, that:

$$\begin{aligned} (++)^i \oplus (+)^i \oplus (-)^i &= ++^i \oplus (-)^i = ++^{i,-i}; \\ ++^i \oplus (+)^i \oplus (-)^i &= ++^i \oplus ? = ? \end{aligned}$$

Heuristics, such as separately adding all positive and all negative signs, must be designed to prevent unnecessary ambiguous results from sign-addition.

The operators revisited We have seen that sign addition can result in signs with a list of multiplication indices. To provide for adding signs with lists of multiplication indices, the \oplus -table needs to be generalised to situations where $i = i_1, \dots, i_n$ and $j = j_1, \dots, j_m$. The generalisation is quite straightforward, only the combinations a) through d), describing the combination of a strong and a conflicting weak influence, deserve some special attention.

We extend the constraint a) for the sum of $++^i$ and $-^j$ in Table 4.2. We now make the following observations from Table 4.2:

- negative multiplication indices can only result from the combination of a strong sign having multiplication index i , with a conflicting weak sign having multiplication index $j \geq i$;
- the combination of a strong sign with an arbitrary other sign never results in a weak sign.

From these observations we observe that weak signs cannot have negative multiplication indices. When a strongly positive sign is added to a weakly negative sign, therefore, we must check whether the positive multiplication indices of the strong sign outweigh the multiplication indices of the weak sign.

Suppose i and j are lists of multiplication indices i_1, \dots, i_n , respectively j_1, \dots, j_m such that $i_1 \leq \dots \leq i_n$ and $j_1 \leq \dots \leq j_m$. Then, we can add the weakly negative sign with multiplication index list j to the strongly positive sign with multiplication index list i , if the strongly positive influence is definitely stronger than the weakly negative influence. This is ensured if

- the strongly positive influence is greater than zero, that is, if

$$\sum_k i_k > 0;$$

- for every index in the list j of (positive) indices, there is a smaller positive index in i , or more formally, there exists an index $i_k > 0$ in i , such that

$$i_k \leq j_1, i_{k+1} \leq j_2, \dots, i_{k+m-1} \leq j_m.$$

The constraints b) through d) in Table 4.2 are extended analogously.

We will illustrate adding signs with lists of multiplication indices with some examples.

- $++^{1,2} \oplus -^{1,3} = ++^{1,2,-1,-3}$, since $1 \leq 1$ and $2 \leq 3$
- $++^{1,2} \oplus -^{1,1} = ?$, since $1 \leq 1$ but then $2 > 1$
- $++^{-2,1} \oplus -^2 = ?$, since $-2 + 1 \leq 0$
- $++^{-2,1,2} \oplus -^2 = ++^{-2,1,2,2}$, since $-2 + 1 + 2 > 0$ and $1 \leq 2$

We would like to note that as multiplication indices represent powers of the cut-off value, lists of multiplication indices can often be simplified. For example, $++^{1,2,-1,-3}$ can be simplified to $++^{2,-3}$, since an influence with sign $++^{1,2,-1,-3}$ lies between $\alpha + \alpha^2 - \alpha - \alpha^3$ and 1 and the term in α equals $\alpha^2 - \alpha^3$.

Now that we have extended the definition of the enhanced \oplus -operator to handle lists of multiplication indices, we concentrate again on the \otimes -operator. In Table 4.1 we used single multiplication indices i and j . Building upon the meaning of the indices in terms of cut-off values, the table can be easily generalised to situations where i and j are lists of multiplication indices. For example, suppose we have influences $S^{++i_1, i_2}(A, B)$ and $S^{++j_1}(B, C)$, that is, we have

$$\Pr(b \mid ax) - \Pr(b \mid \bar{a}x) \geq \alpha^{i_1} + \alpha^{i_2} \quad \text{and} \quad \Pr(c \mid by) - \Pr(c \mid \bar{b}y) \geq \alpha^{j_1}$$

for all combinations of values x and y for the set of node B 's parents other than A and node C 's parents other than B , respectively. We then find that

$$\Pr(c \mid axy) - \Pr(c \mid \bar{a}xy) \geq (\alpha^{i_1} + \alpha^{i_2}) \cdot \alpha^{j_1} = \alpha^{i_1} \cdot \alpha^{j_1} + \alpha^{i_2} \cdot \alpha^{j_1} = \alpha^{i_1+j_1} + \alpha^{i_2+j_1}.$$

Thus, the sign of the transitive influence of A on C is $++^{i_1+j_1, i_2+j_1}$. Similar observations apply to the other sign-products in Table 4.1.

The property of symmetry

The basic sign-propagation algorithm for inference with a regular qualitative network explicitly builds on the properties of symmetry, transitivity, and parallel composition of influences. So far, we have addressed the \otimes - and \oplus -operators and have thereby guaranteed the transitivity and parallel-composition properties of influences in an enhanced qualitative network. We now focus on the property of symmetry of influences in an enhanced network. In a regular qualitative probabilistic network, the property of symmetry guarantees that, if a node A exerts an influence on a node B , then node B exerts an influence of the same sign on node A . In an enhanced qualitative network, as in a regular qualitative network, an influence and its reverse are both positive or both negative. The symmetry property, however, does not hold with regard to the relative strengths of an influence and its reverse: the reverse of a strongly positive qualitative influence, for example, may be a weakly positive influence, and vice versa. One way of ensuring that during inference in an enhanced qualitative network signs can be propagated in both directions of an arc, is to specify the signs of all reversed influences can be specified explicitly; these signs will then have to be elicited from the domain experts involved in the network's construction. An alternative way is to introduce an additional sign: as the relative strength of the reverse of an influence is unknown, the reverse is taken to have an ambiguous strength. The reverse of a positive influence would then be ambiguously positive, denoted by the new sign '+?'. The \otimes - and \oplus -tables can be easily extended to incorporate this additional sign; for details the reader is referred to [102]. Much useful information is however lost using these ambiguous signs and we therefore opt for explicitly specifying the signs of reverse influences.

4.2.3 Probabilistic inference revisited

In the previous section, we have shown that the properties of transitivity and parallel composition of influences hold in an enhanced qualitative probabilistic network. The property of symmetry holds for the basic signs, but not with respect to the strengths of influences. In essence there is therefore no property of symmetry. We argued however that such a property can be easily

enforced by specifying two influences per arc. The basic sign-propagation algorithm from Section 3.2 is therefore generalised straightforwardly to apply to enhanced qualitative networks: instead of the regular \otimes - and \oplus -operators, the enhanced operators are now used for propagating and combining signs. Observations have to be entered as either $++^0$ or $--^0$ instead of $+$ or $-$ to prevent unnecessary loss of information during the first multiplication of the sign of the observation and the sign of the influence it traverses. We illustrate the application of the algorithm by means of our running example.

Example 4.20 We consider once again the qualitative *Radiotherapy* network from Figure 4.5 in Section 4.2.1. We begin by illustrating the application of the basic sign-propagation algorithm. Suppose that we enter the sign $+$ for node T . Node T propagates this sign towards node S . Node S receives the sign $+\otimes+=+$. It thereupon computes the sign $+\otimes=-$ and sends it to node L . Node L does not pass on a sign to node R , since the trail from T via L to R is blocked. Node T also sends a positive sign to node R , which passes it on to node L . Node L therefore in addition receives the sign $+\otimes+=+$. The two signs that enter node L are combined and result in the ambiguous sign $-\oplus+=?$. Note that the ambiguous sign arises from the trade-off represented for node L .

Now, consider the enhanced *Radiotherapy* network from Figure 4.6. We enter the sign $++^0$ for node T , reflecting the positive observation for T . We once again apply the sign-propagation algorithm, this time using our enhanced operators. Recall that initially all influences associated with arcs in the digraph have signs with a multiplication-index of 1. Node T computes the sign $++^0\otimes+=+^1$ and sends it towards node S . Node S in turn computes the sign $+^1\otimes=-^1=-^2$ and sends it to node L . Node L therefore receives the sign $-^2$. Node T also computes the sign $++^0\otimes++^1=++^1$ and sends it to node R . Node R thereupon computes the sign $++^1\otimes++^1=++^2$ and passes it on to node L . Node L thus receives the additional sign $++^2$. Combining the two signs that enter node L results in the sign $++^2\oplus-^2=++^{2,-2}$ indicating that the net result is positive. Note that, while in the regular qualitative network the represented trade-off cannot be resolved and results in an ambiguous sign, the trade-off is resolved in the enhanced qualitative probabilistic network. \square

4.2.4 Discussion

To provide for trade-off resolution, we enhanced the formalism of qualitative probabilistic networks by distinguishing between strong and weak influences. We enhanced the multiplication and addition operators to guarantee the transitivity and parallel-composition properties of influences. To handle the asymmetry of an influence's strength we have proposed specifying two influences for each arc. With these enhancements we have generalised the basic sign-propagation algorithm to apply to enhanced qualitative networks. We have shown that our formalism provides for resolving at least some trade-offs, in a qualitative way, that is, without having to resort to numerical computation.

To provide for trade-off resolution in qualitative probabilistic networks, we have added additional signs and augmented the signs with multiplication indices. Now, probabilistic inference in an enhanced qualitative network with many parallel trails may result in signs with long lists of multiplication indices. As it is hard to interpret the meaning of such lists of indices, it is not our

intention to output the augmented signs: the multiplication indices are merely used internally for trade-off resolution. The output, as in a regular qualitative network, is a basic sign for each node that indicates whether the net influence of an observation is positive or negative.

When the sign-propagation algorithm is used with the enhanced operators, it becomes less efficient than with the standard operators. Although a node's enhanced sign can change at most three times, the multiplication indices associated with the sign may change each time the node is visited. Therefore, the sign of a node may change as many times as the number of simple active trails from the node to the observed node. The runtime of the algorithm on an enhanced network is therefore of the order of the number of active simple trails emanating from the observed node in the network's digraph; for dense graphs this can be exponential in the number of nodes. It is more efficient to use the enhanced formalism only for those parts of the network where trade-offs are encountered during inference; this will be especially effective for networks with sparse graphs that model only few trade-offs. How to identify the trade-offs in a network will be discussed in Section 4.4. Another advantage of such local computation is that it requires only local specification of enhanced signs. During the elicitation of signs, domain experts then only have to compare differences in strengths for small sets of influences. As correctly specifying strengths will be harder for experts than correctly specifying the basic sign for an influence, local specification of enhanced signs will make the resulting signs less prone to error. An additional level of relative strength can be added by introducing, for example, '+ + +' and '- - -' signs with an extra cut-off value. This, however, would require domain experts to be able to distinguish between three levels of relative strength and it would render the necessary \otimes - and \oplus -operators more complicated.

As with regular qualitative influences, we have specified the sign of an enhanced influence of a node A on a node B to be independent of other nodes X in the network. As a consequence, comparable to non-monotonic influences that embed a positive influence for one combination of values for X and a negative influence for another combination, a qualitative influence can, for example, be weakly positive for one combination of values for X and strongly positive for another combination of values, given a specific cut-off value. Such an influence is neither strongly positive nor weakly positive and in the current formalism another cut-off value has to be used. In the next section on context-specific influences we will address this issue.

In enhancing qualitative probabilistic networks, we have so far focused on the distinction between strong and weak influences. We will now briefly discuss further extending the formalism with strong and weak synergies and intercausal influences. We recall from Section 4.1 that additive synergies are used for resolving non-monotonicity of influences. A distinction between strong and weak additive synergies is necessary for distinguishing between strong and weak influences after resolving such non-monotonicities in an enhanced network. As an example, we consider a node C with two parents A and B . Suppose that the influence of node B on node C is non-monotonic with provoker A and suppose that A and B exhibit a positive additive synergy on C , that is,

$$\Pr(c \mid abx) + \Pr(c \mid a\bar{b}x) \geq \Pr(c \mid \bar{a}bx) + \Pr(c \mid \bar{a}\bar{b}x),$$

for all combinations of values x for the set X of parents of C other than A and B . We can write this inequality in terms of the influences of B on C for the different values of provoker A :

$$(\Pr(c \mid abx) - \Pr(c \mid \bar{a}bx)) + (\Pr(c \mid a\bar{b}x) - \Pr(c \mid \bar{a}\bar{b}x)) \geq 0.$$

If both these influences are weak, that is, smaller than the cut-off value α , then the above equation is positive, but smaller than 2α . If one influence is strong and one is weak, then the equation yields a value between α and $1 + \alpha$. If both influences are strong, the result lies between 2α and 2. Note that these intervals overlap, since $1 + \alpha \geq 2\alpha$ because $a \leq 1$. As a consequence, we can only ensure that a resolved influence is weak if both possible influences are weak and the equation yields a value smaller than α ; a resolved influence is ensured to be strong if both possible influences are strong and this value is greater than or equal to $1 + \alpha$. Building upon this observation, a possible definition for a *weakly positive additive synergy* is

$$Y^+(\{A, B\}, C) \iff \forall x \ 0 \leq \Pr(c \mid abx) + \Pr(c \mid \bar{a}\bar{b}x) - \Pr(c \mid \bar{a}bx) - \Pr(c \mid a\bar{b}x) \leq \alpha,$$

for all value combinations x for the set X . A *strongly positive additive synergy* is then defined as

$$Y^{++}(\{A, B\}, C) \iff \forall x \ \Pr(c \mid abx) + \Pr(c \mid \bar{a}\bar{b}x) - \Pr(c \mid \bar{a}bx) - \Pr(c \mid a\bar{b}x) \geq 1 + \alpha,$$

for all value combinations x for the set X . Additive synergies that are neither weak nor strong do not provide for resolving non-monotonicity as far as the relative strength of the resulting influence is concerned. A possible solution to provide for extracting at least the basic sign of the resulting influence is to introduce the aforementioned ‘+?’ for additive synergies.

We recall from Chapter 3 that a product synergy associated with a head-to-head node describes the sign of the intercausal influence induced between two of its parents upon observation of that node. At first thought, a reasonable definition for, for example, a strongly positive product synergy of type I exhibited by nodes A and B with respect to the value c_0 of their common child C , would be

$$X^{++}(\{A, B\}, c_0) \iff \Pr(c_0 \mid ab) \cdot \Pr(c_0 \mid \bar{a}\bar{b}) - \Pr(c_0 \mid \bar{a}b) \cdot \Pr(c_0 \mid a\bar{b}) \geq \alpha.$$

Now, the following relation holds between product synergy I and the intercausal influence it induces:

$$\begin{aligned} & \Pr(c_0 \mid ab) \cdot \Pr(c_0 \mid \bar{a}\bar{b}) - \Pr(c_0 \mid \bar{a}b) \cdot \Pr(c_0 \mid a\bar{b}) \\ &= (\Pr(b \mid ac_0) - \Pr(b \mid \bar{a}c_0)) \cdot \frac{\Pr(c_0 \mid a) \cdot \Pr(c_0 \mid \bar{a})}{\Pr(b) \cdot \Pr(\bar{b})}. \end{aligned}$$

As we cannot determine whether the fraction in this equation is greater or smaller than α , the proposed definition for a strong product synergy of type I does not necessarily describe a strong intercausal influence between nodes A and B . From this observation, we conclude that the relative strength of an intercausal influence cannot be derived from the associated product synergy. We observe however that an intercausal influence between two nodes is just a regular qualitative influence. A *strongly positive intercausal influence* can therefore be defined just as a strongly positive influence; *strongly negative intercausal influences* and weak intercausal influences are defined analogously. Instead of specifying product synergies, a domain expert will now have to specify signs of intercausal influences for all possible observations of their common children.

4.3 Context-specific sign-propagation

Probabilistic networks provide, by means of a digraph, for a qualitative representation of the conditional independence relation that is embedded in a joint probability distribution. The digraph captures independences between nodes only, that is, the digraph models independences regardless of values. Additional independences that hold only for certain values of nodes are captured by the conditional probabilities associated with the nodes in the network. A qualitative probabilistic network captures only the independences portrayed by the digraph: information about additional independences for specific combinations of values of nodes is lost. As these additional independences are qualitative in nature, we should be able to capture them in a qualitative network.

For probabilistic networks a notion of *context-specific independence* has been formalised [9, 137] to denote independences that hold only for certain values of nodes. Context-specific independence can be exploited to speed up probabilistic inference as it allows further decomposition of conditional probabilities resulting in a finer-grained factorisation of the joint probability distribution. Context-specific independence occurs often enough that some well-known tools for the construction of probabilistic networks have incorporated special mechanisms to allow the user to more easily specify the conditional probability distributions for the nodes involved [9].

Motivated by these observations, we introduce a notion of context-specific independence, and of context-specific signs more in general, for qualitative networks. We extend the basic formalism of qualitative networks by providing for the inclusion of context-specific information for influences and show that exploiting this information upon inference can prevent unnecessarily weak results. In addition, we show that context-specific information can be incorporated in enhanced qualitative probabilistic networks, as discussed in Section 4.2, as well, thereby providing for a notion of context-specificity of strengths of influences.

4.3.1 Context-independent signs

We have argued that context-specific independences cannot be expressed in a qualitative network's structure. These independences are in fact hidden in the qualitative influences in the network: if the influence of a node A on a node B is positive for one combination of values for B 's other parents X and zero for all other combinations of values of X , then the influence is modelled as a positive influence and the embedded zero influences remain hidden. Note that zero influences are hidden due to the fact that the inequality in the definition of qualitative influence is not strict. We present an example illustrating hidden zeroes.

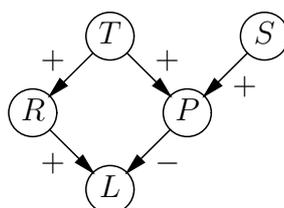


Figure 4.11: The qualitative *surgey* network.

Example 4.21 Figure 4.11 represents a highly simplified and adapted fragment of the prognostic part of the oesophagus network. Node L models the life expectancy of a patient after therapy, where l indicates that the patient will survive for at least one year. Node T models the therapy instilled; here we consider surgery, modelled by t , and no treatment, modelled by \bar{t} , as the only options. The effect to be attained from surgery is a radical resection of the primary tumour, modelled by node R . The most life-threatening condition after surgery is a pulmonary complication, modelled by node P ; the occurrence of this complication is heavily influenced by whether or not the patient under consideration is a smoker, which is modelled by node S .

The probability of attaining a radical resection upon surgery, that is $\Pr(r \mid t)$, equals 0.45; if no surgery is performed, there can be no radical resection and, hence, $\Pr(r \mid \bar{t}) = 0$. From these probabilities we have that node T exerts a positive qualitative influence on node R . The probabilities of a pulmonary complication occurring are given in the following table:

$\Pr(p)$	s	\bar{s}
t	0.75	0.00
\bar{t}	0.00	0.00

From these probabilities, we have that both T and S exert a positive influence on node P . The fact that, for example, the influence of node T on P is actually zero in the context of the value \bar{s} for node S is not apparent from the influence's sign: in the qualitative abstraction of the original probabilistic network this information is lost. Note that the zero influence is not caused by the zero probabilities themselves, but rather by the zero difference between these probabilities.

For the sake of completeness, we also specify the conditional probabilities for life-expectancy given either value of R and P :

$\Pr(l)$	p	\bar{p}
r	0.15	0.95
\bar{r}	0.03	0.50

Node R exerts a positive influence on L , whereas the influence of P on L is negative. Note that, for example, the influence of P on L is quite strong, although this is not apparent from the qualitative signs. \square

From the example above it is apparent that the high level of abstraction provided by qualitative probabilistic networks can cause loss of information and, as a result, may unnecessarily lead to uninformative answers upon probabilistic inference. For example, if a patient is not a smoker, we know that performing surgery has a positive influence on his life expectancy; due to the two conflicting trails from node T to node L , however, entering the observation t for node T will result in the '?'-sign for node L .

We recall that the definition of qualitative influence requires that the sign of an influence from a node A on a node B is independent of the values of the other parents X of B . As a consequence, the initial '?' of a non-monotonic influence hides the information that node A has a positive influence on node B for some combination of values of X and a negative influence for another combination. In Section 4.1, we resolved this problem by explicitly specifying the fact that the influence was non-monotonic and by specifying the nodes provoking the non-monotonicity; we then used the sign of the additive synergy involved to determine the sign of

the resolved influence for a combination of values for the provoking nodes. Recall that the use of additive synergies does not always serve to unambiguously determine the sign of an influence whose non-monotonicity is resolved. Another option is to look upon the non-monotonic influence as specifying different signs for the influence for different contexts. We briefly recapture the *Cervical metastases* example from Section 4.1 to illustrate this observation.

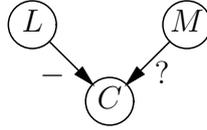


Figure 4.12: The qualitative *Cervical metastases* network.

Example 4.22 Figure 4.12 represents the previously discussed *Cervical metastases* network. The probabilities specified for the presence of cervical metastases in a patient given values for both L and M are:

$\Pr(c)$	l	\bar{l}
m	0.35	0.95
\bar{m}	0.00	1.00

We recall that node L has a negative influence on node C ; the influence of node M on C is non-monotonic:

$$\Pr(c \mid ml) > \Pr(c \mid \bar{m}l), \text{ but } \Pr(c \mid m\bar{l}) < \Pr(c \mid \bar{m}\bar{l}).$$

Node L is the provoker of the non-monotonicity of the influence of node M on node C . An observation of node L will therefore resolve the non-monotonicity, upon which the sign of the resolved influence can be determined. Another way of looking upon the non-monotonic influence is as hiding a ‘+’ for the value l of provoker L and a ‘-’ for the value \bar{l} of L . \square

From the two examples we see that information about zero influences and non-monotonicities that is present in the conditional probability distributions of a probabilistic network, is lost upon abstracting the network in terms of the basic qualitative signs. In the remainder of Section 4.3, we will show that context-specific signs can help to restore some of this information without having to resort to numerical probabilities.

4.3.2 Exploiting context-specific information

The high level of abstraction imposed by qualitative probabilistic networks enforces qualitative influences to be context-independent. As a result, we cannot encode context-dependent information. In this section we present a refinement of the formalism of qualitative networks that allows for associating context-specific signs with qualitative influences.

Context-specific signs

Before introducing context-specific signs, we formally define the notion of context.

Definition 4.23 Let $G = (V(G), A(G))$ be an acyclic digraph. Let $X \subseteq V(G)$ be a set of nodes in G called context nodes. A context c_X for X is a combination of values for a set of nodes $Y \subseteq X$; when $Y = \emptyset$ we say that the context is empty, written ϵ_X . The set of all possible contexts for X is called the context set for X and is denoted by \mathcal{C}_X .

The subscript X for the empty context ϵ will often be omitted when no confusion is possible.

To be able to compare different contexts for the same set of context nodes, we define an ordering on contexts.

Definition 4.24 Let $G = (V(G), A(G))$ be an acyclic digraph and let $X \subseteq V(G)$ be a set of context nodes. Let c_X and c'_X be combinations of values for $Y \subseteq X$ and $Y' \subseteq X$, respectively. Then, $c_X > c'_X$ iff $Y \supset Y'$ and c_X and c'_X specify the same combination of values for Y' .

A context-specific sign δ is a sign that may vary from context to context and can thus be looked upon as a function $\delta : \mathcal{C}_X \rightarrow \{+, -, 0, ?\}$ from contexts for a set of nodes X to signs.

Definition 4.25 Let $Q = (G, \Delta)$ be a qualitative probabilistic network and let $X \subseteq V(G)$. A context-specific sign $\delta(X)$ is a function $\delta : \mathcal{C}_X \rightarrow \{+, -, 0, ?\}$ with the following constraint for any two contexts c_X and c'_X with $c_X > c'_X$:

$$\delta(c'_X) = \delta_i, \delta_i \in \{+, -, 0\} \implies \delta(c_X) \in \{\delta_i, 0\}.$$

To avoid an abundance of braces, we will write $\delta(A)$ instead of $\delta(\{A\})$ to indicate a context-specific sign for a single context node A .

We associate context-specific signs with qualitative influences. A qualitative influence whose sign is context-specific will be called a context-specific influence.

Definition 4.26 Let $G = (V(G), A(G))$ be an acyclic digraph and let Pr be a joint probability distribution such that Pr is an I-map for G . Let A, B be nodes in G with $A \rightarrow B \in A(G)$ and let $X \subseteq \pi_G(B) \setminus \{A\}$. Then, node A exerts a qualitative influence of sign $\delta(X)$ on node B , denoted $S^{\delta(X)}(A, B)$, iff for each context c_X for X we have

- $\delta(c_X) = +$ iff $\text{Pr}(b \mid ac_Xy) \geq \text{Pr}(b \mid \bar{a}c_Xy)$ for any combination of values c_Xy for X ;
- $\delta(c_X) = -$ iff $\text{Pr}(b \mid ac_Xy) \leq \text{Pr}(b \mid \bar{a}c_Xy)$ for any such combination of values c_Xy ;
- $\delta(c_X) = 0$ iff $\text{Pr}(b \mid ac_Xy) = \text{Pr}(b \mid \bar{a}c_Xy)$ for any such combination of values c_Xy ;
- $\delta(c_X) = ?$, otherwise.

We defined a context-specific influence for an arc between two nodes A and B only with respect to context nodes that belong to the set of parents of B . This restriction of the set of context nodes is not essential and can be lifted whenever desirable.

A context-specific sign $\delta(X)$ has to specify a basic sign from $\{+, -, 0, ?\}$ for each possible combination of values in the context set. From the constraint that the sign $\delta(X)$ has to adhere to,

however, we have that it is not necessary for a domain expert to explicitly indicate a basic sign for each combination. For example, suppose that for the influence of a node A on a node B the set of context nodes X consists of nodes D and E . Further suppose that $\delta(X)$ is defined as

$$\begin{aligned} \delta(\epsilon) = ?, \quad \delta(d) = +, \quad \delta(\bar{d}) = -, \quad \delta(e) = ?, \quad \delta(\bar{e}) = +, \\ \delta(de) = +, \quad \delta(d\bar{e}) = +, \quad \delta(\bar{d}e) = -, \quad \delta(\bar{d}\bar{e}) = 0. \end{aligned}$$

It suffices to specify this function only for the smaller contexts, if the larger contexts have the same sign; the function is therefore specified by

$$\delta(\epsilon) = ?, \quad \delta(d) = +, \quad \delta(\bar{d}) = -, \quad \delta(\bar{e}) = +, \quad \delta(\bar{d}\bar{e}) = 0.$$

Note that the sign $\delta(X)$ describes a non-monotonic influence.

The concept of context-specific qualitative influence extends straightforwardly to intercausal influences by providing for a context-specific product synergy.

Probabilistic inference revisited

As we will show shortly, the use of context-specific signs allows us to exploit additional information hidden by regular qualitative influences, whenever an appropriate context is observed. The basic sign-propagation algorithm for probabilistic inference with a qualitative network is easily adapted to apply to both influences with regular and context-specific signs. The adaptation of the sign-propagation algorithm is similar to the one we proposed for propagating over non-monotonic influences in Section 4.1: before propagating a sign over an influence, it is investigated whether or not the influence's sign is context-specific. If the sign is context-specific, we determine the appropriate context from the observed nodes and propagate the sign specified for this context. If no observations are available for the context nodes, then the sign specified for the empty context is used. The adapted algorithm is given in Figure 4.13.

```

procedure PropagateSign(trail, from, to, messagesign):
    sign[to] ← sign[to] ⊕ messagesign;
    trail ← trail ∪ {to};
    for each active neighbour  $V_i$  of to
    do linksign ← sign of (induced) influence between to and  $V_i$ ;
        if linksign ≡  $\delta(X)$ 
        then linksign ←  $\delta(c_X)$  for observations  $c_X$ ;
        messagesign ← sign[to] ⊗ linksign;
        if  $V_i \notin \text{trail}$  and sign[ $V_i$ ] ≠ sign[ $V_i$ ] ⊕ messagesign
        then PropagateSign(trail, to, Vi, messagesign)

```

Figure 4.13: The extended sign-propagation algorithm for handling context-specific signs.

Context-independent versus context-dependent signs

In Section 4.3.1 we argued that context-specific independences can be hidden in the signs of a qualitative network's influences. Revealing these hidden independences and exploiting them during probabilistic inference can be worthwhile. First of all, the information that an influence is zero for a certain context can be used to improve the running time of the sign-propagation algorithm because propagation of a message can be stopped once a zero influence is encountered. Secondly, the context information can help to resolve trade-offs during inference and will thereby forestall unnecessarily weak results.

Example 4.27 We reconsider the *Surgery* network from Figure 4.11. Suppose that a patient is undergoing a surgical removal of his oesophagus. Applying the basic sign-propagation algorithm after entering the observation t into the network results in the sign '?' for node L : we do not have enough information to resolve the trade-off among the two conflicting trails from node T to node L . We now extend the *Surgery* network with the context-specific sign $\delta(S)$ for the influence of T on P , which is defined by

$$\delta(s) = +, \delta(\bar{s}) = 0 \quad \delta(\epsilon) = +.$$

That is, we have included the additional information that non-smoking patients are not at risk of suffering from pulmonary complications. The thus extended network is shown in Figure 4.14. Now, suppose that the patient undergoing surgery is a non-smoker. Sign-propagation of a '+'

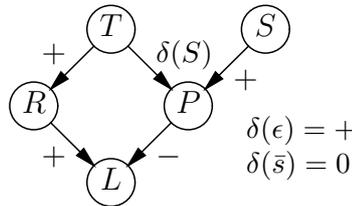


Figure 4.14: A hidden zero revealed by a context-specific sign.

for node T in the context of the observation \bar{s} with the adapted algorithm from Figure 4.13, now results in the sign $(+ \otimes +) \oplus (0 \otimes -) = +$ for node L , that is, we find that surgery is likely to increase the life expectancy for this patient. \square

The example demonstrates that the formalism of context-specific influences provides for revealing hidden zeroes in an elegant way.

In Sections 4.1 and 4.3.1, we discussed that for non-monotonic influences initial '?'s are specified. We argued that it is important to resolve non-monotonic influences to prevent the spreading of '?'s throughout the network during inference. Now that we have extended the formalism of qualitative networks to deal with context-specific information, we have a more direct way of dealing with initially specified '?'s.

Example 4.28 We reconsider the *Cervical metastases* network from Figure 4.12 and illustrate how the non-monotonicity involved can be captured by a context-specific sign. We recall that the influence of node M on node C is non-monotonic, as is apparent from

$$\Pr(c \mid ml) > \Pr(c \mid \bar{m}l), \text{ and } \Pr(c \mid m\bar{l}) < \Pr(c \mid \bar{m}\bar{l}).$$

These inequalities indicate a positive influence of M on C for context l , whereas the influence is negative for the context \bar{l} . Our enriched formalism allows us to capture the influence of node M on node C by the sign $\delta(L)$, which is defined by

$$\delta(l) = +, \delta(\bar{l}) = -, \delta(\epsilon) = ?.$$

The thus enriched network is depicted in Figure 4.15. □

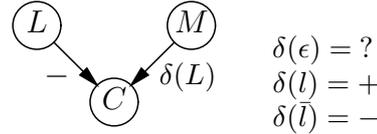


Figure 4.15: A non-monotonicity captured by a context-specific sign.

4.3.3 Extension to enhanced qualitative networks

In the formalism of enhanced qualitative probabilistic networks introduced in Section 4.2, a distinction is made between strong and weak influences. To this end, the set of all influences is partitioned into two disjoint sets of influences using a cut-off value α . Our notion of context-specific sign for regular qualitative probabilistic networks can be easily extended to apply to enhanced qualitative probabilistic networks. To distinguish between regular positive and negative signs and weakly positive and negative signs, we use ‘+?’ and ‘-?’ to denote the former. A context-specific sign for a set of context nodes X then is a function $\delta : \mathcal{C}_X \rightarrow \{++, +?, +, -, -?, --, 0, ?\}$ with the constraints that a strongly positive sign for a context c'_X must be strongly positive for all larger contexts c_X , a weakly positive sign for a context c'_X must be weakly positive or zero for all larger contexts, and a regular positive sign for c'_X may be strongly positive, weakly positive, or zero for larger contexts; similar constraints apply to negative signs. Context-specific signs are once again associated with influences, as in regular networks.

We recall that for distinguishing between strong and weak qualitative influences in an enhanced network, we have to choose a cut-off value α such that for *all* strong influences of a node A on a node B we have

$$|\Pr(b \mid ax) - \Pr(b \mid \bar{a}x)| \geq \alpha,$$

for *all* combinations of values x for the set of node B 's parents other than node A , and for *all* weak influences we have

$$|\Pr(b \mid ax) - \Pr(b \mid \bar{a}x)| \leq \alpha,$$

for *all* such combinations of values x . If, for a specific cut-off value, there exists an influence of node A on node B such that $|\Pr(b \mid ax) - \Pr(b \mid \bar{a}x)| > \alpha$ and $|\Pr(b \mid ax') - \Pr(b \mid \bar{a}x')| < \alpha$ for some combinations of values x and x' , then a different cut-off value must be chosen: α must be shifted towards 0 or 1 and may even end up being 0 or 1. The use of context-specific signs for enhanced influences can prevent this shifting to be necessary, as is illustrated in the following example.

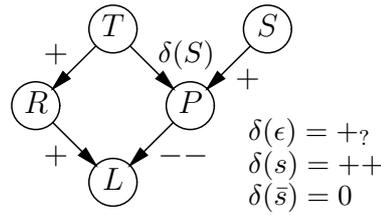


Figure 4.16: Context-specific signs in enhanced qualitative networks.

Example 4.29 We consider once again the example network from Figure 4.11. In the enhanced qualitative network formalism we would like to distinguish between strong and weak influences. Choosing a cut-off value of, for example, $\alpha = 0.46$, we can model the fact that pulmonary complications strongly influence life expectancy, that is, we can specify $S_G^{--}(P, L)$. For this cut-off value, however, the influence of node T on node S is neither strongly positive nor weakly positive; in fact, the influence is strongly positive for the value s of node S and zero for \bar{s} . The cut-off value $\alpha = 0.46$, therefore, does not serve to partition the set of influences in two distinct subsets. To ensure that all influences in the network are either strong or weak, the cut-off value should be either 0 or 1. Using the context-specific sign $\delta(S)$ defined by

$$\delta(s) = ++, \delta(\bar{s}) = 0, \delta(\epsilon) = +?$$

for the influence of node T on node P , we can now explicitly specify the otherwise hidden strong and zero influence. The thus extended network is shown in Figure 4.16.

In Example 4.27 we saw that for non-smoking patients the effect of surgery on life expectancy is positive. For smokers, the effect could not be determined. Using the distinction between strong and weak signs, we can now determine the effect of surgery on life expectancy for smokers to be negative: upon propagating the observation t for node T in the context of the information s for node S , the sign

$$(+^1 \otimes +^1) \oplus (+ +^1 \otimes - -^1) = -^2, -^2$$

results for node L . □

4.3.4 Discussion

In this section we have introduced the notion of context-specific signs for qualitative probabilistic networks. By doing so, we have provided for a finer level of representation detail than in regular qualitative networks. In regular networks, an influence between two nodes can only be unambiguous if it has the same sign regardless of the values of any other nodes in the network, that is, if the influence is context-independent. As a consequence, context-specific information, that is, context-specific ‘+’s, ‘-’s and ‘0’-s, remain hidden. Our extension allows us to explicitly specify such context-specific signs. We have shown that exploiting this information can forestall unnecessary ambiguous node signs during inference. Incorporating the notion of context-specificity into enhanced qualitative probabilistic networks renders even more expressive power. The fact that zeroes and double signs can now be specified context-specifically allows them to be specified more often, in general. We have shown that these zeroes and double signs can be very powerful for resolving trade-offs.

We have extended the formalism of qualitative probabilistic networks with the concept of context-specific signs for influences and product synergies. The concept of context-specificity can also be extended to additive synergies. The concept of context-specific additive synergy, however, is not very interesting as additive synergies are only used to determine the sign of a resolved non-monotonic influence. We recall from Section 4.1 that only the provoking nodes for a non-monotonic influence had to be specified; observations for the provokers then served to resolve the non-monotonicity, using the signs of the appropriate additive synergies to determine the sign of the resolved influence. If the concept of context-specific sign were to be used, not only the provokers, that is, the context nodes, of a non-monotonic influence had to be specified, but also the sign of the resolved influence for the different combinations of values for the provokers. From the context-specific signs specified, the sign of the influence after resolution follows immediately and additive synergies are not required. The use of context-specific signs for determining the sign of a resolved non-monotonic influence should not be automatically preferred to the use of additive synergies. Although, with two or more provokers, additive synergies do not always serve to establish the sign of a resolved non-monotonic influence, it may be easier for domain experts to indicate provokers and specify the synergistic effects between nodes, than it is to specify signs for *combinations of values* of nodes.

Recall that the notion of context-specific independence was introduced for probabilistic networks as a concept that can be exploited to speed up probabilistic inference. Generally, a probabilistic network's conditional probability distributions have to be inspected to determine the presence of context-specific independences [9]. An additional advantage of using context-specific signs in qualitative networks is that, once the network is quantified, context-specific independence information is readily available.

Finally, although our presentation has focused on binary nodes only, we have made no assumptions that would disallow the use of context-specific signs in networks including non-binary nodes.

4.4 Pivotal pruning of trade-offs

In the previous two sections we have detailed two refinements of the formalism of qualitative probabilistic networks that provide for resolving trade-offs during probabilistic inference. With these refinements, it is still possible trade-offs remain unresolved, giving rise to an uninformative result upon inference. In this section we propose a different approach to dealing with trade-offs. We propose to isolate the unresolved trade-offs and identify from the network the *information* that would serve to resolve them, rather than resolving them by providing an even finer level of detail. We present an algorithm for dealing with unresolved trade-offs that builds upon the idea of zooming in on the part of a qualitative probabilistic network where the actual trade-offs reside. After a new observation is entered into the network, probabilistic inference will provide the sign of the influence of this observation on a node of interest, given previously entered observations. If this sign is ambiguous, then there are trade-offs present in the network. In fact, a trade-off must reside along the active trails, or reasoning chains, between the observation and the node of interest. Our algorithm now isolates these reasoning chains to constitute the part of the network that is relevant for addressing the trade-offs present. From this relevant part, an informative result

is constructed for the node of interest in terms of values for the nodes involved and the relative strengths of the influences between them.

4.4.1 Outline of the algorithm

If a qualitative probabilistic network models one or more trade-offs, it will typically yield ambiguous results upon inference with the basic sign-propagation algorithm. Once an ambiguous sign is introduced, it will spread throughout most of the network and an ambiguous sign is likely to result for a specific node of interest. By zooming in on the part of the network where the actual trade-offs reside and identifying the information that would serve to resolve these trade-offs, a more insightful result can be constructed. We illustrate the basic idea of our algorithm to this end.

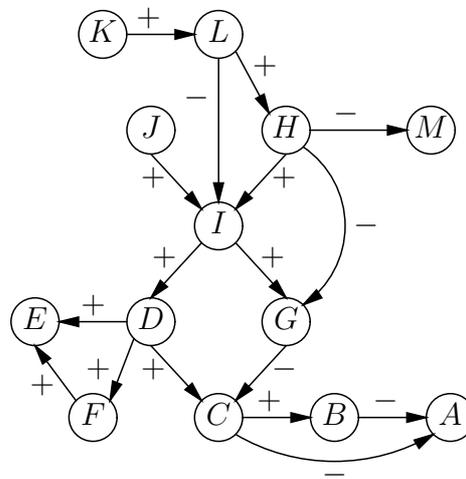


Figure 4.17: The example qualitative probabilistic network.

As our running example, we consider the qualitative probabilistic network from Figure 4.17; for ease of exposition, only the qualitative influences in the network are shown. Suppose that the value *true* is observed for node *H* and that we are interested in its influence on the probability distribution of node *A*. Tracing the influence of the sign ‘+’ for the observation for node *H* on every other node’s distribution by means of the sign-propagation algorithm, results in the node signs shown in Figure 4.18. The ‘?’-sign for node *A* reveals that at least one trade-off must reside along the reasoning chains between the observed node *H* and the node of interest *A*. These reasoning chains together constitute the part of the network that is relevant for addressing the trade-offs that gave rise to the ambiguous result for node *A*; we call this part the *relevant network*. For the example, the relevant network is shown in Figure 4.19 below the dashed line. Our algorithm isolates this relevant network for further investigation by deleting from the network all nodes and arcs that are connected to, but no part of the reasoning chains from *H* to *A*.

A relevant network for addressing trade-offs typically includes many nodes with ambiguous node signs. Often, however, only a small number of these nodes are actually involved in the trade-offs that gave rise to the ambiguous result for the node of interest. Figures 4.18 and 4.19, for example, reveal that, while the nodes *A*, *B*, and *C* have ambiguous node signs, the influences

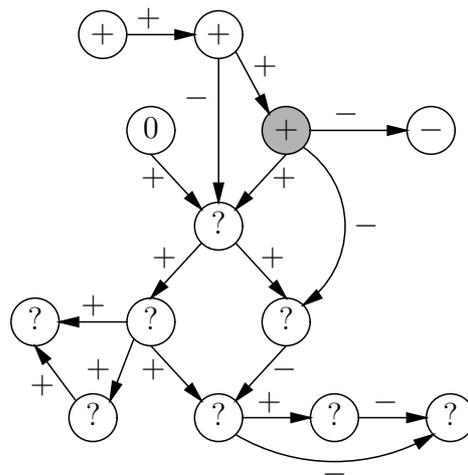


Figure 4.18: The result of propagating '+' for node H .

between them are not conflicting. In fact, every possible unambiguous node sign $sign[C]$ for node C would result in the unambiguous sign $sign[C] \otimes ((+ \otimes -) \oplus -) = sign[C] \otimes -$ for node A . For addressing the trade-offs involved, therefore, the part of the relevant network between node C and node A can be disregarded. Node C in fact separates the part of the relevant network that contains trade-offs from the part that does not. We call node C the *pivot node* for the node of interest.

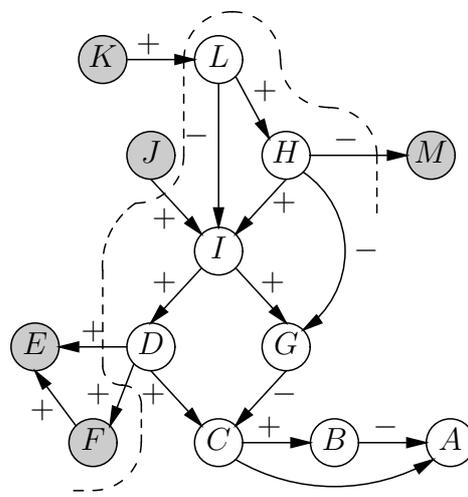


Figure 4.19: The relevant network, below the dashed line.

In general, the pivot node in a relevant network is a node with an ambiguous node sign for which every possible unambiguous sign would uniquely determine an unambiguous sign for the node of interest; in addition, no other node having this property resides on an active trail from the observed node to the pivot node, that is, the pivot node is the node with this property “closest” to the observed node. Note that the node of interest may itself obey the properties of a pivot node; every network therefore includes a pivot node. Our algorithm now selects from the relevant network the pivot node for the node of interest.

From the definition of pivot node, it can be shown that there must be two or more different active trails in the relevant network from the observed node to the pivot node; the net influences along these trails, moreover, must be conflicting or ambiguous. To resolve the ambiguity at the pivot node, the relative strengths of the various influences as well as the signs of some of the nodes involved need be known. From Figures 4.18 and 4.19, for example, we have that node I lies at the basis of the ambiguous sign for the pivot node C . Note that it receives an ambiguous node sign itself as a result of two conflicting (non-ambiguous) influences. An unambiguous node sign for node I , however, would not suffice to fix an unambiguous sign for node C . Even knowledge of the relative strengths of the two conflicting influences from node I on the pivot node would not suffice for this purpose: a positive node sign for node I , for example, would still cause node G , residing on one of the active trails from I to C , to receive an ambiguous node sign, which in turn gives rise to an ambiguous influence on C . Node G therefore also lies at the basis of the ambiguity at the pivot node. Every combination of unambiguous node signs for the nodes G and I would render the separate influences on the pivot node unambiguous: knowledge of the relative strengths of these influences would suffice to determine an unambiguous sign for the pivot node. We call a minimal set of nodes having this property the *resolution frontier* for the pivot node.

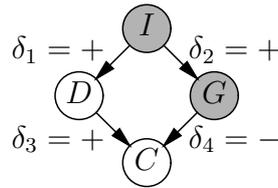


Figure 4.20: The construction of a sign for node C .

In terms of signs for the nodes from the resolution frontier, our algorithm constructs a (conditional) sign for the pivot node by comparing the relative strengths of the various influences exerted on it upon inference. In the example network, the nodes from the resolution frontier exert two separate influences on the pivot node C : the indirect influence from node I via node D on C and the direct influence from G on C . For the sign δ of the influence of node I via node D on C we find that

$$\delta = \text{sign}[I] \otimes \delta_1 \otimes \delta_3 = \text{sign}[I] \otimes +$$

and for the sign δ' of the influence of G on C we find that

$$\delta' = \text{sign}[G] \otimes \delta_4 = \text{sign}[G] \otimes -,$$

where δ_i , $i = 1, 3, 4$, are as in Figure 4.20. For the node sign $\text{sign}[C]$ of the pivot node, the algorithm now constructs and reports the following result:

$$\text{if } |\delta| \geq |\delta'|, \text{ then } \text{sign}[C] = \delta, \text{ else } \text{sign}[C] = \delta';$$

where $|\delta|$ denotes the strength of the sign δ . So, if the two influences on node C have opposite signs, then their relative strengths will determine the sign for node C . The sign of the node of interest A then follows directly from the node sign of C .

4.4.2 Splitting up and constructing signs

In this section we detail some of the issues involved in our algorithm for pivotal pruning of trade-offs. In doing so, we assume that a qualitative probabilistic network does not include any ambiguous influences, that is, ambiguous node signs upon inference result only from unresolved trade-offs. We further assume that observations are entered into the network one at a time and that sign propagation resulted in an ambiguous sign for the network's node of interest. For ease of reference, Figure 4.21 summarises the pivotal-pruning algorithm in pseudocode.

```

procedure PivotalPruning( $Q$ ):
     $Q_{rel} \leftarrow \text{ComputeRelevantNetwork}(Q)$ ;
     $pivot \leftarrow \text{ComputePivot}(Q_{rel})$ ;
    ConstructResults( $Q_{rel}, pivot$ )
  
```

Figure 4.21: The basic algorithm for pruning trade-offs.

In detailing the algorithm, we focus attention on identifying the relevant part of a qualitative probabilistic network along with its pivot node and on constructing from these an informative result for the node of interest.

Identifying the relevant network

Our algorithm identifies from a qualitative probabilistic network the relevant part for addressing the trade-offs that resulted in an ambiguous sign for the node of interest. We begin by formally defining the concept of relevant network.

Definition 4.30 *Let $Q = (G, \Delta)$ be a qualitative probabilistic network. Let $O \subseteq V(G)$ be the set of previously observed nodes, let $E \in V(G)$ be the node for which new evidence has become available, and let $I \in V(G)$ be the network's node of interest. The relevant network for E and I given O is the qualitative probabilistic network $Q_{rel} = (G', \Delta')$ such that*

- $V(G')$ consists of all nodes that occur on an active trail from E to I ;
- $A(G') = (V(G') \times V(G')) \cap A(G)$; and
- Δ' consists of all influences and synergies from Δ involving nodes from G' only.

Various notions of relevance have been introduced, most notably for quantitative probabilistic networks [35, 112]. We briefly review some of these notions and illustrate their differences. For a node of interest I , previously observed nodes O , and a newly observed node E , we say that a node N is

- *structurally relevant* to I , if N is not d-separated from I given $O \cup \{E\}$;
- *computationally relevant* to I , if the (conditional) probabilities for N are required for computing the posterior probability distribution for I given the observations for $O \cup \{E\}$; and

- *dynamically relevant* to I and E , if N partakes in the impact of E on I in the presence of the observations for O .

In the example qualitative network from Figure 4.17, node D is structurally relevant, computationally relevant, and dynamically relevant to the node of interest A . Node E is structurally relevant to node A , yet neither computationally nor dynamically relevant. Node J is structurally irrelevant to the observed node H , as is also evidenced by its node sign ‘0’ upon inference; it is both structurally and computationally relevant to the node of interest A , yet dynamically irrelevant. The newly observed node H is d-separated from A by its being observed. It therefore is not structurally relevant to A ; it is computationally as well as dynamically relevant to A , however. Node M , to conclude, is neither structurally nor computationally or dynamically relevant to the node of interest A .

The concept of *dynamic relevance* was introduced to denote all nodes constituting the reasoning chains between a newly observed node and a node of interest in a probabilistic network [35]. The concept of dynamic relevance closely resembles our concept of relevance; in fact, the set of all nodes in a network’s digraph G that are dynamically relevant to the node of interest I and the newly observed node E , given the previously observed nodes O , induces the digraph G' of the relevant network for E and I given O as defined in Definition 4.30. To identify the relevant network for a newly observed node E and a node of interest I given the previously observed nodes O , therefore, it is sufficient to compute the nodes that are dynamically relevant to E and I . The dynamically relevant nodes are identified by first determining all nodes that are computationally relevant to the node of interest I and then removing the *nuisance* nodes that are not on any reasoning chain from the newly observed node E to I [35]. For computing the set of all computationally relevant nodes, the efficient *Bayes-Ball* algorithm is available from R.D. Shachter (1998). The algorithm takes for its input a probabilistic network, the set of all observed nodes $O \cup \{E\}$, and the node of interest I ; it returns, among other things, the sets of nodes that are computationally relevant, or *requisite*, to I . From the set of computationally relevant nodes, all nuisance nodes for E and I , that is, all nodes that are not on any reasoning chain from the newly observed node E to the node of interest I , need to be identified. An efficient algorithm is available for identifying these nodes [75]. The algorithm takes for its input a computationally relevant network, the set of previously observed nodes O , the newly observed node E , and the node of interest I ; it returns the set of nuisance nodes for E and I . The algorithm for computing the relevant part of a qualitative probabilistic network is summarised in pseudocode in Figure 4.22.

Identifying the pivot node

After establishing the relevant part of a qualitative probabilistic network for addressing the trade-offs present, our algorithm identifies the pivot node. We recall that the pivot node serves to separate the part of the relevant network that contains the trade-offs that gave rise to the ambiguous sign for the node of interest, from the part that does not contain such trade-offs. The pivot node will thus allow for further focusing. We define the concept of pivot node more formally.

Definition 4.31 *Let $Q = (G, \Delta)$ be a relevant qualitative probabilistic network for the newly observed node E and the node of interest I , given the previously observed nodes O . Then, the pivot node for I and E is a node $P \in V(G) \setminus O$ for which we have that*

```

function ComputeRelevantNetwork( $Q$ ):  $Q$ 
   $requisites \leftarrow \text{BayesBall}(G, O \cup \{E\}, I)$ ;
   $V(G) \leftarrow (V(G) \setminus requisites) \cup \{E\}$ ;
   $A(G) \leftarrow (V(G) \times V(G)) \cap A(G)$ ;
   $nuisances \leftarrow \text{ComputeNuisanceNodes}(G)$ ;
   $V(G) \leftarrow V(G) \setminus nuisances$ ;
   $A(G) \leftarrow (V(G) \times V(G)) \cap A(G)$ ;
   $\Delta \leftarrow \{\text{all influences and synergies from } \Delta \text{ in } G\}$ ;
  return  $Q_{rel} = (G, \Delta)$ 

```

Figure 4.22: The algorithm for computing the relevant network.

- $\hat{S}^?(E, P, t_i) \in \Delta$, where t_i is the parallel trail composition of all simple trails in G from E to P ;
- $\hat{S}^\delta(P, I, t_j) \in \Delta$ for some $\delta \neq '?'$, where t_j is the parallel trail composition of all simple trails in G from P to I ; and
- no node P' with the above properties exists that resides on an active trail from E to P .

Note that in the previous definition the composite trail composed of t_i and t_j comprises all simple trails from node E to node I . We now have that the pivot node is a node with an ambiguous node sign for which every possible unambiguous sign would uniquely determine an unambiguous sign for the node of interest.

The pivot node in a relevant qualitative probabilistic network has various convenient properties that allow for its easy identification. One of these properties is that the pivot node is shared by all active trails from the observed node to the node of interest. If we assume that intercausal influences induced by previous observations are added to the network's digraph as undirected edges, then the graph's articulation nodes have the same property. To see this, we recall from Chapter 2 that an articulation node is a node that upon removal along with its incident arcs, makes the digraph fall apart into various separate components. In the digraph of our example network, as shown in Figure 4.17, the articulation nodes are the nodes C , D , H , I , and L . For the relevant network, depicted in Figure 4.19, node C is the only articulation node; node C also happens to be the pivot node. Note that, in general, the node of interest cannot be an articulation node in a relevant network, whereas it can be the pivot node. Propositions 4.32 now states that the pivot node is either an articulation node or the node of interest.

Proposition 4.32 *Let $Q = (G, \Delta)$ be a relevant qualitative probabilistic network for the newly observed node E and the node of interest I , given the previously observed nodes O . The pivot node for I and E is either the node of interest I or an articulation node in G .*

Proof: From the definition of pivot node, we have that every possible unambiguous node sign for the pivot node determines an unambiguous sign for the node of interest I . It will be evident that node I itself satisfies this property. Either the node of interest I or another node on an active trail from E to I , therefore, is the pivot node. Now, suppose that node I is not the pivot node.

As a sign for the pivot node uniquely determines the sign for I , we conclude that all influences exerted upon I must traverse the pivot node. Every active trail from E to I , therefore, must include the pivot node. As the relevant network consists of only active trails from E to I and the pivot node is an unobserved node, removing the pivot node along with its incident arcs will cause the network to fall apart into separate components. We conclude that the pivot node is an articulation node of the relevant network. \square

As the pivot node is the node with an unambiguous influence on the node of interest closest to the newly observed node, the pivot node is unique.

Proposition 4.33 *Let $Q = (G, \Delta)$ be a relevant qualitative probabilistic network for the newly observed node E and the node of interest I , given the previously observed nodes O . The pivot node for I and E is unique.*

Proof: From Definition 4.30 we have that the relevant network consists of only nodes that reside on an active trail from the newly observed node E to the node of interest I . From the definition of articulation node we further have that every such trail must include all articulation nodes in the relevant network. In fact, every active trail from E to I visits the articulation nodes in the same order. From the last condition in Definition 4.31 we have that no pivot node can reside on the active trail from another pivot node to the node of interest. We conclude that the pivot node is unique. \square

Articulation nodes are identified using a depth-first search algorithm; for details and an algorithm, we refer the reader to [50]. For identifying the pivot node we now observe that the articulation nodes in a relevant network allow a total ordering, as already indicated in the proof of Proposition 4.33. We number the articulation nodes, together with the node of interest I , from 1, for the node closest to the newly observed node, to m , for the node of interest. The pivot node then is the node with the lowest ordering number for which an unambiguous sign would uniquely determine an unambiguous sign for the node of interest. Our algorithm for determining the pivot node starts with investigating the articulation node closest to the node of interest; this node is numbered $m - 1$. The algorithm investigates whether an unambiguous sign for this candidate pivot node would result in an unambiguous sign for the node of interest upon sign propagation. By propagating a '+' from the candidate pivot node to the node of interest I , the node sign resulting for I is the sign of the net influence of the candidate pivot node on I . If this sign is ambiguous, then the node of interest itself is the pivot node. Otherwise, the algorithm proceeds by investigating the articulation node numbered $m - 2$. This process is continued until the articulation node i is found such that node i has an unambiguous influence on the node of interest and node $i - 1$ has an ambiguous influence on the node of interest. The algorithm is summarised in pseudocode in Figure 4.23.

Constructing results

From its definition, we have that there must be two or more different reasoning chains in the relevant network from the newly observed node to the pivot node; the net influences along these reasoning chains are conflicting or ambiguous. Our algorithm focuses on the ambiguity at the

```

function ComputePivot( $Q$ ): node

   $candidates \leftarrow \{I\} \cup \text{FindArticulationNodes}(G)$ ;
  order the nodes from  $candidates$  from 1 to  $m$ ;
  return FindPivot( $m - 1$ );

function FindPivot( $i$ ): node

  PropagateSign( $\emptyset$ , node  $i$ , node  $i$ , ‘+’)
  if sign[node  $i + 1$ ] = ‘?’
  then return node  $i + 1$ ;
  else FindPivot( $i - 1$ )

```

Figure 4.23: The algorithm for computing the pivot node.

pivot node and identifies the information that would serve to resolve it. For this purpose, the algorithm zooms in on the part of the relevant network between the newly observed node and the pivot node; we call this part the *pruned relevant network*. The pruned relevant network consists of all active trails between the newly observed node E and the pivot node P from the relevant network. Note that the pruned relevant network is readily computed by exploiting the property that the pivot node is an articulation node. From the pruned relevant network, the algorithm first selects the so-called *candidate resolvers*.

Definition 4.34 Let $Q = (G, \Delta)$ be a relevant qualitative probabilistic network for the newly observed node E and the node of interest I , given the previously observed nodes O . Let P be the pivot node for I and E . Now, let $Q_{pru} = (G', \Delta')$ be the pruned relevant network for P . A candidate resolver for P is a node $R_i \in V(G')$, $R_i \neq P$, such that

- $R_i = E$, or
- $sign[R_i] = \text{'?'}$ and $in-degree_{G'}[R_i] \geq 2$.

The candidate resolvers for the pivot node are easily identified from the pruned relevant network. In our example network, as shown in Figure 4.19, the candidate resolvers for the pivot node C are the nodes H , I , and G .

From among the candidate resolvers in the pruned relevant network, our algorithm now constructs the resolution frontier. We recall that the resolution frontier is a minimal set of nodes for which unambiguous node signs would uniquely determine the signs of the separate influences on the pivot node.

Definition 4.35 Let $Q = (G, \Delta)$ be a pruned relevant qualitative probabilistic network for the pivot node P , the newly observed node E and the node of interest I , given the previously observed nodes O . Let R be the set of candidate resolvers for P . The resolution frontier F for P is the maximal subset of R , with respect to set inclusion, such that for each candidate resolver $R_i \in F$ there exists an active trail from node E via node R_i to P for which no other candidate resolver $R_j \in R$ resides on the subtrail from R_i to P .

In Figure 4.19, the resolution frontier for the pivot node C consists of the nodes I and G .

Proposition 4.36 states that the resolution frontier for a pivot node is unique.

Proposition 4.36 *Let $Q = (G, \Delta)$ be a pruned relevant qualitative probabilistic network for pivot node P , the newly observed node E and node of interest I , given the previously observed nodes O . The resolution frontier F for P is unique.*

Proof: Suppose that we have two different resolution frontiers F and F' for P . We now show that this assumption leads to a contradiction. From the definition of resolution frontier we have that for each node $F_i \in F$ there exists an active trail from E via F_i to P for which no other candidate resolver resides on the subtrail from F_i to P ; a similar observation holds for each node in F' . For each $F_i \in F \cup F'$ then, the same property holds. From F and F' being resolution frontiers, we have that both sets are maximal with respect to set inclusion. However, the set $F \cup F'$ is a larger subset of the candidate resolvers that obeys the properties of a resolution frontier for P . This contradicts the assumption that F and F' are resolution frontiers. From the contradiction we conclude that the resolution frontier for the pivot node is unique. \square

The resolution frontier for a pivot node can be easily constructed by recursively traversing the different active trails from the pivot node back to the observed node E and checking whether the visited nodes are candidate resolvers. As soon as a candidate resolver is found on an active trail, the traversal of the trail is halted.

Once the resolution frontier is computed from the pruned relevant network, the algorithm constructs a (conditional) sign for the pivot node in terms of signs for the nodes from the frontier. Let F be the resolution frontier for the pivot node P . For each resolver $R_i \in F$, let s_j^i , $j \geq 1$, denote the signs of the different active trails from R_i to the pivot node. For each combination of node signs $sign[R_i]$, $R_i \in F$, the sign of the pivot node is reported to be

$$\text{if } \left| \bigoplus_{(sign[R_i] \otimes s_j^i)=+} (sign[R_i] \otimes s_j^i) \right| \geq \left| \bigoplus_{(sign[R_i] \otimes s_j^i)=-} (sign[R_i] \otimes s_j^i) \right| \\ \text{then } sign[P] = +, \text{ else } sign[P] = -, \quad (4.4)$$

where $|\delta|$ once again is used to denote the strength of the sign δ . We would like to note that as, in general, the resolution frontier includes a small number of nodes, the number of signs to be computed for the pivot node is limited. In addition, we note that the process of constructing informative results can be repeated recursively for the nodes in the pivot node's resolution frontier, until the newly observed node is reached. The basic algorithm for constructing results is summarised in pseudocode in Figure 4.24.

Time complexity of the algorithm

Our algorithm for pruning a qualitative probabilistic network for trade-offs, basically consists of three steps: computing the relevant network, computing the pivot node, and constructing results.

- The relevant network $Q_{rel} = (G', \Delta')$ is computed from a qualitative network $Q = (G, \Delta)$ using the BayesBall algorithm and the algorithm for identifying nuisance nodes. Both these algorithms have a time complexity of $O(|V(G)| + |A(G)|)$.

```

procedure ConstructResults( $Q, pivot$ ):
     $Q_{pru} \leftarrow \text{ComputePrunedNetwork}(Q, pivot)$ ;
     $candidates \leftarrow \text{ComputeCandidates}(Q_{pru}, pivot)$ ;
    output ComputeResults( $Q_{pru}, pivot, candidates$ )

function ComputeResults( $Q_{pru}, pivot, candidates$ ): text
     $frontier \leftarrow \text{ComputeFrontier}(pivot, \emptyset, candidates)$ ;
    for all  $R_i \in frontier$ 
    do determine  $s_j^i, j \geq 1$ ;
    for all  $R_i \in frontier$  and  $sign[R_i] = +, -$ 
    do return statement (4.4);

function ComputeFrontier( $pivot, frontier, candidates$ ): set of nodes
    for all  $V_i$  such that  $V_i \rightarrow pivot \in A(G_{pru})$  or  $pivot \rightarrow V_i \in A(G_{pru})$ 
    do if  $V_i \in candidates$ 
        then  $frontier \leftarrow frontier \cup \{V_i\}$ 
        else ComputeFrontier( $V_i, frontier, candidates$ )

```

Figure 4.24: The algorithm for constructing results.

- The pivot node is determined by computing all articulation nodes in the relevant network and performing a number of sign propagations. Finding the articulation nodes takes $O(|V(G')| + |A(G')|)$ time; sign-propagation is done for each articulation node at most and therefore takes $O(m \cdot |A(G')|)$, where m is the number of articulation nodes.
- Constructing the results amounts to computing the pruned relevant network $Q_{pru} = (G'', \Delta'')$ and from that the candidate resolvers and the resolution frontier. For the nodes in the resolution frontier the final output is computed. The pruned network can be computed in $O(|V(G')| + |A(G')|)$ time. Identifying the candidate resolvers takes $O(|V(G'')|)$ time, whereupon the resolution frontier is determined in $O(c)$ time, where c is the number of candidate resolvers. Finally, for computing the results, we have to determine the sign of influence of each node in the resolution frontier on the pivot node; this takes $O(f \cdot |A(G'')|)$ time, where f is the number of nodes in the resolution frontier.

Under the assumption that the number of articulation nodes and the number of nodes in the resolution frontier is bound by a constant, we conclude that the runtime complexity of the pivotal-pruning algorithm is $O(|V(G)| + |A(G)|)$.

4.4.3 Discussion

We have presented a new algorithm for dealing with trade-offs in qualitative probabilistic networks. Rather than resolve trade-offs by providing for a finer level of representation detail, our

algorithm identifies from a qualitative probabilistic network the *information* that would serve to resolve the trade-offs present. For this purpose, the algorithm zooms in on the part of the network where the actual trade-offs reside and identifies the pivot node for the node of interest. The sign of the pivot node uniquely determines the sign of the node of interest. For the pivot node, a more informative result than ambiguity is reported in terms of values for the node's resolvers and the relative strengths of the influences upon it. This process of constructing informative results can be repeated recursively for the pivot node's resolvers.

We would like to note that for computing informative results for a relevant network's pivot node, the pruned network can be even further restricted. To this end, a so-called *boundary node* can be identified for the newly observed node. The boundary node is the articulation node closest to the node of interest that has an unambiguous node sign after propagation of the new observation. Constructing results can then focus on the part of the relevant network between the pivot node and the boundary node. Moreover, if the thus pruned network includes many articulation nodes, it may very well be that trade-offs exist between the articulation nodes numbered $k - 1$ and k , but not between k and $k + 1$. Distinguishing between these components is straightforward and allows for further focusing on the actual trade-offs involved in inference.

In this section, we have made no assumptions about the type of node, binary or non-binary, in the network. Our pivotal-pruning algorithm can be applied to any qualitative probabilistic network by applying the dummy-value approach introduced in Chapter 3 for sign-propagation.

Our concepts of pivot node and resolution frontier for zooming in on trade-offs and constructing insightful results for a network's node of interest is a very powerful concept to study the reasoning behaviour of qualitative probabilistic networks as well as for enabling explanation of complex reasoning processes in quantitative probabilistic networks. The use of qualitative probabilistic networks for explanation purposes was proposed before [32, 56]. Explanation was explored, for example, using the concept of *knots* to explain the propagation of evidence from knot to knot [118]. Knots are nodes in a dynamically relevant network that are shared by all active trails between an evidence node and a node of interest, with the exception of head-to-head nodes with (indirect) evidence. These knots in fact coincide with the articulation nodes in a network where induced intercausal influences are added as edges. As our algorithm builds on concepts related to knots, it can be seen as an interesting extension of existing work in the field of explanation.

4.5 Propagating multiple simultaneous observations

In the previous sections, we addressed the propagation of a single observation through a qualitative probabilistic network. The signs computed for the nodes in the network indicate the direction of shift in a node's probability distribution occasioned by this single observation, in the light of previously entered observations. In real-life applications, often the simultaneous, joint effect of multiple observations is of interest. Multiple observations can in essence be dealt with by the basic sign-propagation algorithm in two ways [31]. The first approach is to enter and propagate the observations one after the other; the results of the successive propagations then are combined to yield their joint effect. The other approach is to create a single dummy node D and arcs $O_i \rightarrow D$ for each observed node O_i ; the sign of the influence associated with a newly created arc $O_i \rightarrow D$

corresponds to the sign of the observation for node O_i . Running the sign-propagation algorithm with the single observation ‘+’ for the dummy node D will now yield the joint effect of all the observations.

For the first approach, the order in which multiple observations are entered into the qualitative network can influence the net result. The differences originate from the dynamics of the set of influences that are exploited for the propagation of signs: at any time during inference, this set of influences is determined by the network’s digraph and the previously entered observations. Also, depending on the order in which the observations are entered, propagation of multiple observations can yield results that are weaker than necessary, that is, an ambiguous sign results instead of ‘+’, ‘−’ or ‘0’. For the second approach, the node sign of the dummy node is fixed; the node signs of the truly observed nodes on the other hand are not fixed and can therefore change during inference. Again inference can lead to unnecessarily weak results as, for example, a node with a positive observation could now receive a negative sign and update its node sign to a ‘?’-sign. Since upon propagation of multiple observations more influences tend to be combined than for a single observation, more ambiguous signs are likely to result with both approaches.

In this section, we address propagation of multiple observations in a qualitative probabilistic network and focus on the problem of ambiguous signs. We will show that knowledge about the dynamics of the set of influences that is used during sign-propagation of each observation can be exploited to yield stronger and, hence, more informative results upon propagation of multiple observations.

4.5.1 Dynamics of influences and sign-propagation

During probabilistic inference in both qualitative and quantitative probabilistic networks, probabilistic independences between the nodes are exploited using the d-separation criterion. Upon entering an observation, nodes that were d-separated can become dependent given the new evidence and nodes that were not d-separated can become independent [123]. So, trails in the digraph that were once blocked can become unblocked, and vice versa. We give some examples to illustrate the possible effects of these dynamics on the results of propagating multiple observations. We discuss the situation in which blocked trails become unblocked and the opposite situation separately.

Unblocking blocked trails

When using the approach of propagating multiple observations sequentially, the order in which they are entered can affect the net result and yield a result weaker than necessary. The literature on the sign-propagation algorithm is not very explicit in stating whether or not an intercausal influence that is induced by an observation, is used *immediately* upon sign propagation, that is, whether or not an observation for a node is passed on to its parents only through direct influences or also through the intercausal influence just induced. We will show that regardless of whether or not the intercausal influence is used immediately, the order in which observations are entered can affect the net result.

Example 4.37 We consider the *Wall invasion* network from Figure 3.2 in Chapter 3. Suppose that a patient's tumour is longer than 10 *cm* and has grown beyond the oesophageal wall, invading neighbouring structures. We enter the two observations one after the other and investigate the results yielded upon inference.

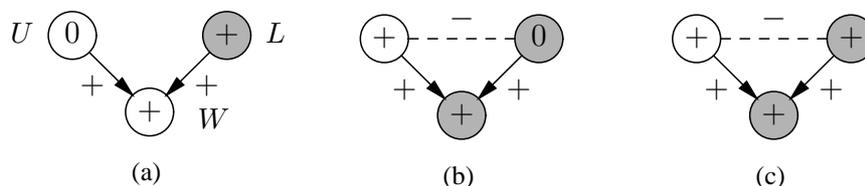


Figure 4.25: The separate effects of subsequently entering a '+' for L , (a), and W , (b), and their joint effect (c).

Figure 4.25 shows the result from first entering the observation that the tumour is longer than 10 *cm* and then the fact that it has grown beyond the oesophageal wall. Upon entering the observation for the tumour's length, the probability that neighbouring structures are invaded increases: after a '+' is entered for node L , it sends a '+' to node W . As ulceration and a tumour's length are independent causes of wall invasion, U 's probability distribution is not affected by the observation. Then, a '+' is entered for node W to indicate the observation of invaded neighbouring structures. Node W sends a '+' to both U and L . The probability of an ulcerating tumour increases. As the tumour's length has been established, L 's probability distribution is not affected by the observation. Note that the intercausal influence induced between U and L does not have any effect, regardless of whether or not it is used immediately. The joint effect of the observations shows an increase in the probability of an ulcerating tumour.

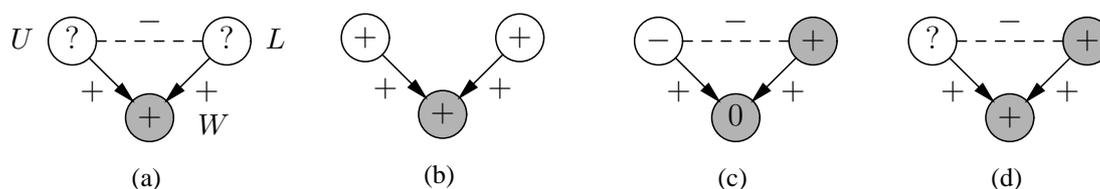


Figure 4.26: The separate effects of entering a '+' for W , exploiting, (a), or disregarding, (b), the intercausal influence, and subsequently entering a '+' for L , (c), and their joint effect, (d).

Figure 4.26 shows the result from first entering the observation that neighbouring structures are invaded and then the observation of the tumour's length. Upon entering the observation for wall invasion, the probabilities of its two causes increase. The observation, however, also induces a negative intercausal influence between U and L . If this intercausal influence is used immediately, U and L propagate a '-' to one another over this influence, resulting in a '?' for both; otherwise both their signs remain positive. Subsequently entering the observation for the tumour's length will now cause the probability of an ulcerating tumour to decrease: node L sends a '-' over the intercausal influence to U . It further sends a '+' to W , but, as wall invasion has been observed, W will not change sign. The joint effect of the observations reveals that the net effect on node U 's probability distribution is unknown. \square

The previous example shows that the order in which multiple observations are entered into a qualitative network can lead to different node signs upon inference; moreover, the node signs yielded can be weaker than necessary. In the example, the difference in results can be attributed to the intercausal influence and the moment it is effectuated. Upon propagating an observation for node L in Figure 4.25, the trail $L \rightarrow W \leftarrow U$ is blocked. In Figure 4.26, however, the blocked trail has become unblocked due to the observation of the head-to-head node W . The set of influences used during the propagation of an observation for node L therefore differs in the two situations.

The induced intercausal influence also leads to different results when using the dummy-node approach for propagating signs, as illustrated by the following example.

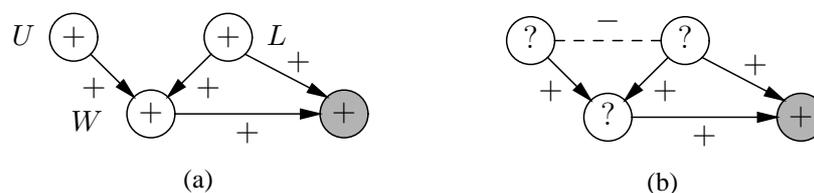


Figure 4.27: Propagating observations using the dummy node approach and disregarding (a), or exploiting (b) the intercausal influence.

Example 4.38 We consider once again the *Wall invasion* network and observations l for node L and w for node W . For propagating these observations, we add a dummy node to the network with incoming arcs from the two observed nodes. The signs along the arcs correspond to the signs of the corresponding observations. A '+' is then entered, as the observation for the dummy node. Figure 4.27 shows the results from propagating with and without exploiting the intercausal influence induced by the observation for the dummy. \square

The example illustrates a typical problem of the dummy-node approach: as the observations for node L and node W are entered as an observation for the dummy node, the signs of node L and node W can change during inference, which may cause additional spreading of '?'s.

Blocking unblocked trails

When using the approach of propagating multiple observations sequentially, the order in which observations are entered influences the order in which trails are blocked. This, in turn, may affect the set of influences over which each subsequent observation is propagated. We again show the effect of entering multiple observations in different orders and of using the dummy-node approach.

Example 4.39 We consider the highly simplified fragment of the oesophagus network from Figure 4.28(a). The fragment again pertains to the invasion of a carcinoma into the oesophageal wall, modelled by node W , and one of its causes, the length of the tumour, modelled by node L . Another effect of the tumour's length is whether or not the tumour is circular, modelled by node

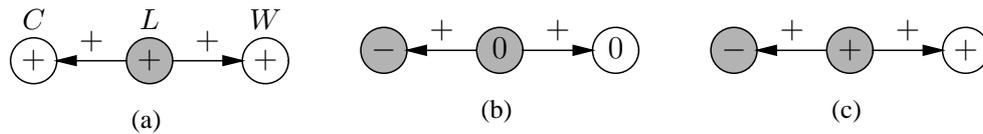


Figure 4.28: The separate effects of subsequently entering a ‘+’ for L (a) a ‘-’ for C (b), and their joint effect (c).

C . The longer a tumour is, the higher the probability that it is a circular tumour. Suppose that a patient’s tumour is longer than 10 cm , but that it is not circular. We enter the two observations one after the other and investigate the results yielded upon inference.

Figure 4.28 shows the result from first entering the observation that the tumour is longer than 10 cm and then the information that it is not circular. Upon entering the observation for the tumour’s length, the probabilities of a circular tumour and of an invasion of neighbouring structures increase. Then, a ‘-’ is entered for node C to indicate the observation that the tumour is not circular. As node L is observed, its sign is not changed; in addition, as node L blocks the trail from node C to node W , it does not pass on any signs to W and the sign of node W remains unaffected. The joint effect of the two observations shows an increase in the probability that the carcinoma has grown beyond the oesophageal wall.

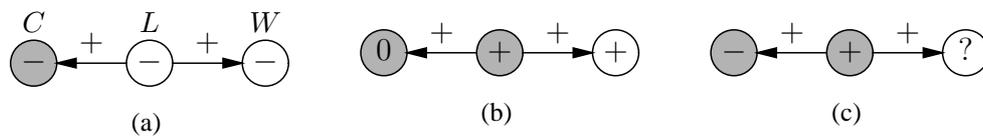


Figure 4.29: The separate effects of subsequently entering a ‘-’ for C (a) a ‘+’ for L (b), and their joint effect (c).

Figure 4.29 shows the result from first entering the observation that the tumour is not circular and then the information that it is longer than 10 cm . The joint effect of the observations reveals that the net effect on node W ’s probability distribution is ambiguous. Finally, Figure 4.30 shows the result from entering the two observations by means of a dummy node.

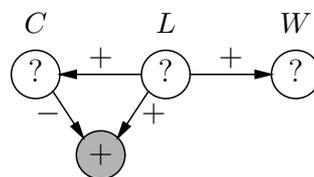


Figure 4.30: The effect of entering a ‘+’ for L and a ‘-’ for C using the dummy node approach.

We would like to note that if we would have had an additional arc $C \rightarrow X$ with a positive influence associated with it, then it would have had a positive sign with the approach of subsequently entering observations, but again a ‘?’-sign with the dummy-node approach. \square

Summary

When propagating multiple observations by sequential propagations of the separate observations and adding the results, the order in which the observations are entered can influence the results. The differences can be attributed to the dynamics of the set of influences over which signs are propagated. By entering observations, influences are removed from this set as a consequence of trails being blocked; also, influences are added to the set as a result of intercausal influences being induced. In Section 4.5.2, we will show that by predicting the dynamics that will be caused by a set of observations before propagating them, we can guarantee sign-propagation of multiple observations to result in the strongest possible sign. In addition, in Section 4.5.3 we will show that direct influences always dominate over intercausal influences, which means that the latter can be disregarded when propagating multiple simultaneous observations. As the dummy node approach requires changes to the network's digraph and, in addition, seems to always result in the weakest signs we will from here on only consider the approach of sequentially propagating observations and combining signs.

4.5.2 Exploiting the dynamics

When a single observation is entered into a qualitative network, its sign is propagated to each node that is not d-separated from the observed node, given previously observed nodes. Suppose we have two simultaneously observed nodes O_1 and O_2 , and no previous observations; we propagate the observations sequentially. If the observation for node O_1 is entered first, its sign is propagated to all nodes that are not d-separated from node O_1 . Subsequently, the sign of the observation for O_2 is propagated to all nodes that are not d-separated from O_2 given O_1 . Similarly, entering the observation for O_1 after the observation for O_2 results in the propagation of O_1 's sign to all nodes that are not d-separated from O_1 given O_2 . To ensure that the order in which these observations are entered has no effect on the outcome, the sign of an observation should be propagated only along trails that will not be blocked by subsequent observations. To this end, for each observed node O_i in a set O of multiple simultaneously observed nodes, we need to determine the nodes that are not d-separated from O_i given all observations for $O_j \in O, i \neq j$.

Definition 4.40 *Let $G = (V(G), A(G))$ be an acyclic digraph. Let $P \subseteq V(G)$ be a set of previously observed nodes and let $O \subseteq V(G) \setminus P$ be the set of multiple simultaneously observed nodes. Then, for each $O_i \in O$, the exclusion set $X(O_i)$ is the set X of all nodes X_k for which $\langle \{O_i\} \mid (O \cup P) \setminus \{O_i\} \mid \{X_k\} \rangle_G^d$ holds.*

The following proposition states that any node in the exclusion set of a node O_i is probabilistically independent of O_i given $(O \cup P) \setminus \{O_i\}$.

Proposition 4.41 *Let G, P , and O be as in the previous definition. Let \Pr be a joint probability distribution on $V(G)$ such that G is an I-map for \Pr . Then, for each $O_i \in O$,*

$$\Pr(X(O_i) \mid O \cup P) = \Pr(X(O_i) \mid (O \cup P) \setminus \{O_i\}).$$

Proof: The proof follows directly from the definition of $X(O_i)$ and Definition 2.25. □

From the above proposition, we have that for propagating observations for a set of nodes O , the propagation of the observation for node $O_i \in O$ should be restricted to all nodes in the digraph that are not included in $X(O_i)$; in doing so, we prevent the order of entering multiple observations to affect the net result. A node's exclusion set can be computed with the Bayes-Ball algorithm mentioned in Section 4.4. The algorithm returns for a probabilistic network with a set of observed nodes and a node of interest, the set of nodes that are computationally relevant for the node of interest given the observed nodes, as well as the set of nodes that are *structurally irrelevant* for the node of interest given the observed nodes [112]. The set of structurally irrelevant nodes for an observed node O given all other observations, is our exclusion set for O .

The idea of propagating observations only to nodes that are not in the exclusion set of the current observation can also be exploited for solving a problem with the dummy node approach. Recall that a problem of the dummy-node approach is that the node signs of observed nodes can change as the observations are entered as an observation for the dummy node. We can remedy this by adapting the sign-propagation algorithm so that it only sends those signs to observed nodes that originate directly from the dummy node.

4.5.3 Dominance of direct over intercausal influences

In the previous section, we showed how to prevent the order in which multiple observations are entered into a qualitative network from affecting the net result of probabilistic inference. In this section, we focus on the intercausal influences that are added to the set of influences as a consequence of entering observations. We show that these intercausal influences can be disregarded during inference, yielding possibly stronger results.

In Section 4.5.1 we mentioned that the current literature does not specify whether or not an induced intercausal influence should be used for propagation of the observation that induced it. We show that, upon propagating multiple observations, we can disregard the intercausal influences induced by these observations.

Proposition 4.42 *Let $Q = (G, \Delta)$ be a qualitative probabilistic network. Let A, B, C be nodes in G such that $A \rightarrow C$ and $B \rightarrow C \in A(G)$. Let node C be observed and let δ_C be the sign of the observation; let δ_A be the sign computed for node A after propagation of the observation for node C . Then,*

$$S^\delta(A, C) \implies \delta_A = \delta_C \otimes \delta.$$

Proof: We will prove the proposition for $\delta = +$ and $\delta_C = +$, that is, we assume that we have observed the value c for node C ; proofs for other combinations of δ and δ_C are analogous. Let $X = \pi_G(C) \setminus \{A, B\}$ be the set of all parents of C other than A and B . Let Pr be a joint probability distribution such that G is an I-map for Pr , then

$$S^+(A, C) \iff \forall x \text{Pr}(c \mid ax) - \text{Pr}(c \mid \bar{a}x) \geq 0.$$

The sign computed for node A captures the change in A 's probability distribution occasioned by the observation for node C and therefore equals the sign of the difference $\text{Pr}(a \mid c) - \text{Pr}(a)$. By

applying Bayes' rule to $\Pr(a | c)$, we find that for all combinations of values x for the set X ,

$$\Pr(a | cx) - \Pr(a) = \frac{\Pr(c | ax) \cdot \Pr(a)}{\Pr(c | x)} - \Pr(a) = \frac{(\Pr(c | ax) - \Pr(c | x)) \cdot \Pr(a)}{\Pr(c | x)}.$$

By conditioning, in the numerator, on A and B , we find that for all x

$$\Pr(a | cx) - \Pr(a) = \frac{\Pr(a) \cdot \Pr(\bar{a})}{\Pr(c | x)} \cdot ((\Pr(c | abx) - \Pr(c | \bar{a}bx)) \cdot \Pr(b) + (\Pr(c | a\bar{b}x) - \Pr(c | \bar{a}\bar{b}x)) \cdot \Pr(\bar{b})).$$

From $S^+(A, C)$ we conclude that $\Pr(a | c) - \Pr(a) \geq 0$ and, hence, that $\delta_A = \delta_C \otimes \delta$. \square

In the proposition we have assumed that an observation had been obtained for a child C of A . The proposition, however, also holds for an indirect observation of node C , that is, for an observation of a descendant D of C . The node sign for node A is then determined by the sign of the observation of node D and the sign of the influence $\hat{S}^\delta(A, D, t)$ along the parallel trail composition of all active trails from A to D .

From the proposition above, we have that the intercausal influence induced by an observation can be disregarded when propagating that observation. We will now show that if the observation pertains to a node from a set of simultaneously observed nodes, then the intercausal influence induced by the observation can also be disregarded when subsequently propagating the other observations; more specifically, we will show that direct influences always dominate over intercausal influences.

Dominance of direct over intercausal influences was already suggested in the Ph.D. thesis of M.J. Druzdzal in Section 6.4.3 [31]. The section focuses on the situation where a head-to-head node and one of its parents are observed. In this situation, the effect of the observation of the head-to-head node on the unobserved parent is larger than the effect of the observation for the other parent via the intercausal influence induced. Druzdzal claims that this dominance property is captured by the following statement:

For parents A, B of C , and A, C of D , the qualitative influence of D on B solely depends on the qualitative influence of D on C and that of C on B .

In the network described, there are two simple trails from D to B , one consisting of $D \leftarrow A \rightarrow C \leftarrow B$ and one consisting of $D \leftarrow C \leftarrow B$. The latter trail is the only active trail, which renders the statement correct. Unfortunately, the statement does not mention observed nodes nor induced intercausal influences and therefore cannot capture the dominance property. Note that if node D , a descendant of the head-to-head node C , is observed, then an intercausal influence is induced between nodes A and C and the trail from A to B via C becomes active. In our opinion, there is no way of qualitatively showing that the (intercausal) influence along this latter trail can be disregarded.

We now prove that the dominance property of direct over intercausal influences indeed holds. We start by formally defining the concept of dominance of one influence over the other.

Definition 4.43 Let $G = (V(G), A(G))$ be an acyclic digraph. Let A, B, C be nodes in G , such that $A \rightarrow C$, and $B \rightarrow C \in A(G)$. Then, the influence of node A on node C is dominated by the influence of node B on C iff, for all observations $a_i \in \{a, \bar{a}\}$ of A and $b_i \in \{b, \bar{b}\}$ of B , we have

$$|\Pr(c | b_i) - \Pr(c)| \geq |\Pr(c | a_i) - \Pr(c)|.$$

The definition states that the influence of node B on node C dominates the influence of node A on C , if propagation of an observation for B has a larger effect on the probability distribution for node C than propagating an observation for A .

We are now interested in whether direct influences dominate over intercausal influences; more specifically, we are interested in the situation where the signs of the messages sent across these influences during sign-propagation are conflicting. The following proposition states that, in this situation, the direct influence dominates over the intercausal influence.

Proposition 4.44 Let $Q = (G, \Delta)$ be a qualitative probabilistic network. Let A, B, C be nodes in G such that $A \rightarrow C, B \rightarrow C \in A(G)$. Let δ_{direct} be the sign of change computed for node A given an observation for node C and let δ_{inter} be the sign of change computed for node A given a subsequent observation for node B . Let δ_A be the node sign computed for node A given both observations. Then,

$$\delta_{direct} = - \otimes \delta_{inter} \implies \delta_A = \delta_{direct}.$$

Proof: We will prove the proposition for $\delta_{direct} = +$; more in particular, we prove the proposition for $S^+(A, C)$ and observation c for node C . We suppose that $Z^-(A, B, c)$. In addition, we suppose that the observation b for B is obtained so that δ_{inter} is negative. Proofs for the other situations are analogous.

The sign of change δ_{direct} in node A 's probability distribution occasioned by the observation for node C equals the sign of the difference $\Pr(a | c) - \Pr(a)$. The sign of change δ_{inter} in node A 's probability distribution occasioned by the subsequent observation for node B equals the sign of $\Pr(a | bc) - \Pr(a | c)$. The node sign δ_A that is computed for node A after propagating both observations equals the sign of the difference

$$\Pr(a | bcx) - \Pr(a) = \Pr(a | cx) - \Pr(a) + \Pr(a | bcx) - \Pr(a | cx),$$

for all combinations of values x for the set $X = \pi_G(C) \setminus \{A, B\}$. By applying Bayes' rule to $\Pr(a | bcx)$, we find that

$$\Pr(a | bcx) - \Pr(a) = \frac{\Pr(c | abx) \cdot \Pr(a | bx)}{\Pr(c | bx)} - \Pr(a).$$

By exploiting independence of A and B and by rearranging terms, we find that

$$\Pr(a | bcx) - \Pr(a) = \frac{(\Pr(c | abx) - \Pr(c | bx)) \cdot \Pr(a)}{\Pr(c | bx)}.$$

By conditioning $\Pr(c | bx)$, in the numerator, on A , we find that

$$\Pr(a | bcx) - \Pr(a) = (\Pr(c | abx) - \Pr(c | \bar{a}bx)) \cdot \frac{\Pr(a) \cdot \Pr(\bar{a})}{\Pr(c | bx)}.$$

From $S^+(A, C)$ we have that $\Pr(c \mid abx) - \Pr(c \mid \bar{a}bx) \geq 0$ for all x and, therefore, for all x $\Pr(a \mid bcx) - \Pr(a) \geq 0$. We conclude that $\delta_A = \delta_{direct}$. \square

The proposition also holds for indirect observations for node C . We note that when the two signs are not conflicting, the two combine into a unique, informative sign. From the dominance property, we now have that during the sequential propagation of multiple simultaneous observations, intercausal influences induced by any of these observations can be disregarded.

4.5.4 Probabilistic inference revisited

In the previous sections, we showed that observations for the nodes O_i from a set O of observed nodes can be propagated only to the nodes that are not d-separated from O_i given all other observed nodes; in addition, we showed that upon propagating multiple simultaneous observations we can disregard the intercausal influences induced by these observations. The basic sign-propagation algorithm can be easily adapted to incorporate these properties. To this end, we first ensure that messages are sent over trails that are active with respect to the previously entered observations; intercausal influences induced by previous observations are thereby exploited. From there on, any new intercausal influence that is induced by the current set of observations is disregarded. In addition, for each subsequent observation propagation is restricted to nodes that are not in the exclusion set of the node to which the observation pertains. The resulting algorithm is summarised in pseudocode in Figure 4.31.

```

procedure PropagateObservations( $Q, Obs, signs, OldObs$ ):
    for each  $V_i$  do  $sign[V_i] \leftarrow '0'$ ;
    for each  $O_i \in Obs$ 
    do  $X(O_i) \leftarrow \text{Bayes-Ball}(G, (Obs \cup OldObs) \setminus O_i, O_i)$ ;
        PropagateSign( $\emptyset, O_i, O_i, sign_i$ )

procedure PropagateSign( $trail, from, to, messagesign$ ):
     $sign[to] \leftarrow sign[to] \oplus messagesign$ ;
     $trail \leftarrow trail \cup \{to\}$ ;
    for each neighbour  $V_i$  of  $to$  that is active given  $from$  and  $OldObs$ 
    do  $linksign \leftarrow \text{sign of (induced) influence between } to \text{ and } V_i$ ;
         $messagesign \leftarrow sign[to] \otimes linksign$ ;
        if  $V_i \notin trail$  and  $V_i \notin X(O_i)$ 
        and  $sign[V_i] \neq sign[V_i] \oplus messagesign$ 
        then PropagateSign( $trail, to, V_i, messagesign$ )

```

Figure 4.31: The adapted sign-propagation algorithm for sequentially propagating multiple simultaneous observations.

To determine the exclusion sets for each observed node, the Bayes-Ball algorithm is executed. This algorithm takes $O(|V(G)| + |A(G)|)$ time for each execution. Sign-propagation of a single

observation takes $O(|A(G)|)$ time. Propagation of multiple simultaneous observations therefore takes $O(m \cdot (|V(G)| + |A(G)|))$ time, where m is the number of observations that is propagated.

We illustrate the impact of disregarding intercausal influences and using exclusion sets upon propagating multiple observations with an example.

Example 4.45 We consider the highly simplified fragment of the oesophagus network shown in Figure 4.32(a). The figure pertains to the diagnostic part of the network. Nodes G_L and G_S model the outcomes of a gastroscopic examination with regard to the location of the tumour in the patient’s oesophagus, modelled by node L , and the macroscopic shape of the tumour, modelled by node S , respectively. The outcomes of this examination depend on how deep the scope can be entered into the patient’s oesophagus; the ease with which the scope can be entered is directly related to how well the patient is able to swallow food, modelled by node P . The shape of the carcinoma influences the possibility of decay of tissue, or necrosis, occurring, modelled by node N ; it further determines the depth of invasion of the tumour into the oesophageal wall, modelled by node W . The depth of invasion in turn has effect on the presence of haematogenous metastases, modelled by node H , which can be found in, for example, the lungs and the liver; the presence of metastases in the lungs and the liver are modelled by the nodes Lu and Li , respectively.

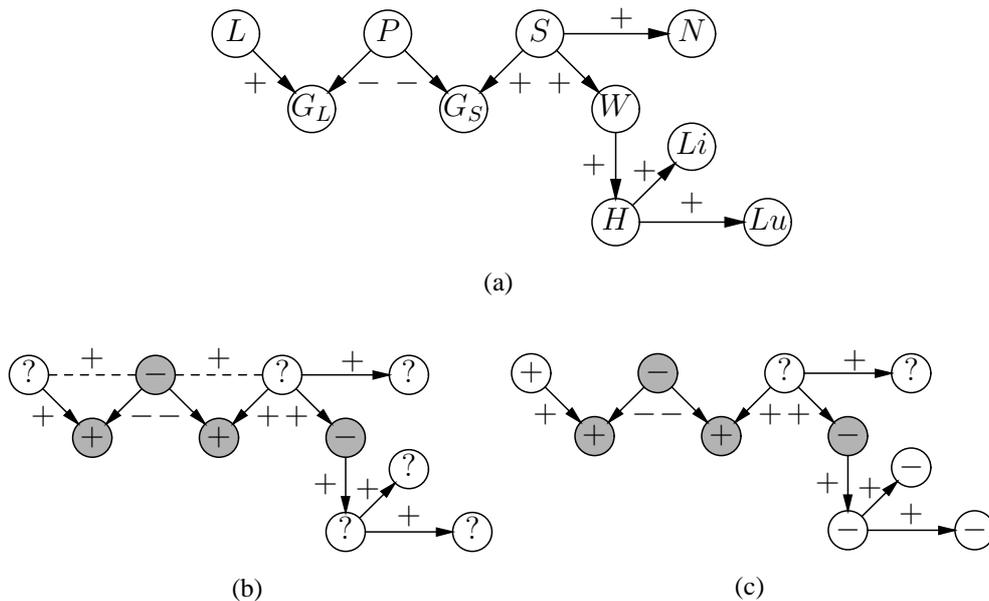


Figure 4.32: Propagation of multiple observations in a fragment of the oesophagus network (a), using the original sign-propagation algorithm (b) and the adapted algorithm (c).

Figure 4.32(b) shows the result of using the original sign-propagation algorithm, after entering the subsequent observations $G_S = true$, $G_L = true$, $P = false$, and $W = false$, and combining the results. Note that the nodes L and S receive a ‘?’ as a result of the negative sign propagated over the intercausal links induced by the observations of G_S and G_L . Since the observation for node W is entered last, it has not blocked the propagation of these ‘?’-signs to H ,

Li and Lu . Figure 4.32(c) shows the result of using the adapted algorithm on the set of observations; the intercausal influences are disregarded and the fact that node W d-separates the set of nodes $\{H, Li, Lu\}$ from the other observed nodes is exploited. \square

4.5.5 Discussion

The basic algorithm for probabilistic inference with a qualitative network was designed to determine the effect of a single observation on the nodes' probability distributions given all previously entered observations. Handling multiple observations by applying the basic algorithm for each observation separately and combining the results into their joint effect, can yield results that are weaker than necessary; furthermore, the results may depend on the order in which the observations are entered. As the cause of this problem, we identified the dynamics of the set of influences over which signs are propagated upon inference. We showed that the intercausal influences that are added to this set as a result of entering observations are always dominated by direct influences and can therefore be disregarded upon inference. Note that the fact that an induced intercausal influence can be disregarded when propagating the observation that has induced it, can be exploited in the original sign-propagation algorithm as well. In addition, we showed that we can prevent propagation of an observation to nodes that will be d-separated from that observed node given subsequent observations, by computing an exclusion set for each observed node. By means of an example, we demonstrated that exploiting these properties yields stronger results, and that the order in which observations are entered no longer influences the net result.

The results presented in this section are not restricted to networks containing binary nodes only; the proofs presented in this section can be easily generalised to networks including non-binary nodes.

4.6 Related work

In this chapter we presented various refinements of the formalism of qualitative probabilistic networks, that aim at preventing ambiguous results upon probabilistic inference as far as possible. We discussed the representation and resolution of non-monotonic influences and we added notions of strength and context to resolve trade-offs modelled in the network. In addition, we presented algorithms for isolating trade-offs and subsequently identifying the information necessary to resolve them, and for propagating multiple simultaneous observations in qualitative probabilistic networks. In this section we provide a brief comparison with related work.

Non-monotonic influences

In [134], M.P. Wellman notes that one of the limitations of the formalism of qualitative probabilistic networks is that they can only express monotonic influences between nodes. To the best of our knowledge, Wellman is the only researcher who has ever referred to the problem of non-monotonic influences and we are the first to propose a solution for it.

S. Parsons addressed a different type of non-monotonicity, in which the influence of a node A on a node B has different signs for different values of node B . To handle this type of non-

monotonicity, Parsons introduces the concept of qualitative derivative, that allows for specifying a sign for each separate value of node B , instead of specifying one sign based upon the cumulative distribution of the values of B [88]. Note that this concept is only of interest for non-binary nodes.

Trade-off resolution

In the context of trade-off resolution, Parsons introduced the concept of categorical influence [88]. A categorical influence is a qualitative influence that serves either to increase a probability to 1 or to decrease a probability to 0, disregarding all other influences. A categorical influence thus serves to resolve any trade-off in which it is involved, but can only capture deterministic relationships between nodes; in real-life applications few to none of such relationships will exist.

Parsons also studied the use of both relative and absolute order-of-magnitude reasoning in the context of qualitative probabilistic networks [88]. Relative orders of magnitude can be used to relate different qualitative influences to each other. Using the relative order-of-magnitude system ROM[K] [27], one qualitative influence can be specified as being *negligible* with respect to, *distant* from, *comparable* to, or *close* to another influence. The use of relative orders of magnitude thus serves to relate the strengths of different influences, but it requires the specification of a relation between pairs of influences, instead of a notion of strength per influence. In addition, the relations used seem to be ill-defined, which makes reasoning with them anything but intuitive. For absolute order-of-magnitude reasoning, Parsons proposes a method that revolves around the propagation of interval probability values. The arcs in a network's graph are labelled as being strongly positive, weakly positive, etc., where a probability interval is associated with each label. Interval comparison is done using \geq_{int} , where $[a, b] \geq_{int} [c, d]$ iff $a \geq c$ and $b \geq d$. Qualitative probabilistic reasoning can be arrived at by just providing the labels for the arcs and not actually quantifying the boundaries of the probability intervals; this approach is comparable to our treatment of the cut-off value introduced in Section 4.2. The method of interval comparison, however, can lead to considerable loss of information.

κ -calculus [29] can be considered another absolute order-of-magnitude system. Using κ -calculus, probabilities can be abstracted to κ -values, where a κ -value of n indicates that the associated probability has the same order of magnitude as ϵ^n for some infinitesimal number ϵ . Drawbacks of the use of κ -calculus [29] are that it is not concerned with *changes* in probabilities, but rather with the probabilities themselves, and that it has been designed for infinitesimal probabilities only.

Categorical influences, order-of-magnitude reasoning and κ -calculus are of a purely qualitative nature, yet serve for resolving only some trade-offs. C.-L. Liu and M.P. Wellman designed methods for resolving trade-offs based upon the idea of reverting to numerical probabilities whenever necessary [76]. The methods presented by Liu and Wellman provide for incrementally applying numeric inference to the point where qualitative reasoning can produce a decisive result. Their methods thereby resolve any trade-off present in the network, but require a fully specified, numerical probabilistic network.

The approaches described above with respect to trade-off resolution have a number of drawbacks; they are either applicable only in networks that model a specific type of relationship, are based on ill-defined relationships, or even require a full numerical specification. In this chapter

we proposed refinements that can easily compete with a number of the above approaches, and can be considered complementary to others.

Trade-off isolation

One of the steps in our algorithm for isolating trade-offs in a qualitative probabilistic network, is the identification of reasoning chains and the nodes these chains have in common. A similar step is used in an algorithm designed for explanation of probabilistic networks [118]. To the best of our knowledge, however, no earlier attempts were made at designing algorithms for isolating trade-offs and subsequently identifying the information required to resolve them.

Propagating multiple observations

The two earlier approaches we discussed for propagating multiple simultaneous observations in a qualitative network, the use of a dummy node and adding the results of subsequent propagations, were proposed by M.J. Druzdzel [31]. Druzdzel, however, does not identify the problems associated with these approaches.

In our discussion of propagating multiple observations, we have taken observations to be only positive or negative and we have not taken into account the fact that one observation can be stronger than another. Parsons, in his discussion of the use of order-of-magnitude approaches, does allow for relating the strengths of multiple observations.

Part II

Probability Elicitation

In which we design and evaluate an elicitation method tailored to fast elicitation of a large number of probabilities. Existing probability elicitation methods are designed to elicit unbiased assessments; these methods are often complicated and time-consuming. Elicitation of the large number of probabilities required for a probabilistic network calls for an elicitation method that accommodates, as much as possible, the experts in the assessment task.

The Elicitation Process

For quantification of a probabilistic network, one often has to rely on domain experts to provide the necessary probabilities. Extensive psychological research has shown that people, even experts, tend to find it difficult to assess probabilities: to simplify the task they use heuristics, most often leading to poorly calibrated and biased assessments [64]. From the field of decision analysis, several methods are available for the elicitation of probabilities [20, 83, 129]. These methods are designed for the elicitation of probabilities in general and not tailored to probability elicitation for probabilistic networks. Most of these methods were designed to overcome, or at least suppress, the problems of bias and poor calibration [64]. However, these methods tend to be so time-consuming that it is infeasible to apply them when hundreds or thousands of probabilities are to be assessed. Faster elicitation methods are available, but are prone to even more biased answers. Before undertaking a large elicitation task, it is therefore important to be aware of the advantages and drawbacks of these methods.

In the field of probabilistic networks, it is well-known that probability elicitation is a problem. We feel, though, that the knowledge about why it is a problem is less wide-spread; it is also less known that there exist various methods, designed especially for probability elicitation. Besides being aware of problems of bias, the builder of a network has to take into consideration not only the method to use, but also, for example, which expert to choose, how to motivate and train the expert, and how to perform the actual elicitation. In this chapter we will give an overview of the entire elicitation process and the available methods, discussing issues to be aware of and to take into consideration when faced with the task of probability elicitation.

This chapter is organised as follows. In Section 5.1 we will first discuss the process of probability elicitation, including motivating and training the expert, the actual elicitation phase, and the verification of the probabilities obtained. Then, in Section 5.2, we will consider the different ways an expert can be presented with the probabilities required and the representation formats that experts can use for indicating their assessment. In Section 5.3 we will discuss various elicita-

tion methods found in, for example, the decision analysis literature, along with their benefits and their drawbacks. As our main concern is probability assessment for probabilistic networks, we will only consider methods for eliciting *discrete* probability distributions. We are interested in point probabilities and will therefore not consider elicitation methods for interval probabilities. Finally, in Section 5.4 we will discuss some matters concerning elicitation methods in general, and draw some conclusions.

5.1 The elicitation process

Research in experimental psychology has shown that simply asking a person to provide a (numerical) probability results in biased probability judgements [64]. To overcome biases, it seems necessary to have a well-structured process for probability elicitation. Such a process is called an *elicitation process* [39, 43, 80]; it can be roughly divided into five stages:

1. select and motivate the expert
2. train the expert
3. structure the questions
4. elicit and document the expert judgements
5. verify the results.

We will further detail these stages in the following subsections, after devoting a subsection to the biases that call for a well-structured elicitation process.

5.1.1 Heuristics and biases

A *bias* is a systematic tendency to take into account factors that are irrelevant to the task at hand, or to ignore relevant facts, thereby failing to make an inference that any appropriate normative theory, for example probability theory, would classify as necessary [42]. Two types of bias can be distinguished: motivational bias and cognitive bias [114]. *Motivational biases* are caused by personal interests and circumstances of the expert. For example, an expert makes careful assessments if he believes that his job depends on the success of the current project; he will be too confident about his assessments, because he, being an expert, feels he should not be uncertain about them. Motivational biases can often be overcome by explaining the expert that an honest assessment is requested, not a promise. *Cognitive biases* arise during the processing of information by the expert and are typically the result of using heuristics [64]. Cognitive biases can be suppressed by informing the expert of their existence and by using different elicitation methods.

When people are asked to make complicated judgements such as probability assessments, they often *subconsciously* use *heuristics*, or rules of thumb, to simplify the task. Four heuristics, among others, are commonly found: availability, anchoring, representativeness, and control [64]. *Availability* is a heuristic with which an expert assesses the probability of an event by the ease

with which occurrences of the event are brought to mind. The idea behind the heuristic is that frequent events are more available, and therefore an event that is easily brought to mind will have a high probability. Often this heuristic works quite well, but it can become a misleading indicator of the frequency with which certain events occur. If, for example, plane crashes are head-line news more often than car crashes, people will *overestimate* the probability of a plane crash and *underestimate* the probability of being involved in a car crash. The process of assessing a probability by choosing an initial value, termed the anchor, and then adjusting up or down from this value, is called the heuristic of *anchoring and adjustment*. Assessments acquired this way are typically biased towards the starting value, due to insufficient adjustment. The resulting bias is termed *anchoring bias*.

The *representativeness* heuristic describes the process where people use the similarity of two events to estimate the degree to which one event is representative of the other. Consider the following well-known example from a study by Tversky and Kahneman [64]:

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Please check off the most likely alternative:

- Linda is a bank teller.
- Linda is a bank teller and is active in the feminist movement.

From this description, most people find it likely that Linda is a feminist and conclude that it is more likely that Linda is a feminist bank teller, than just a bank teller. The example illustrates how a description representative of a feminist can trick people into choosing the less likely event. The cognitive bias, here introduced by the representativeness heuristic, is called the *conjunction fallacy*. A more detailed description seems to be more representative, though the conjunction of two events can never be more likely than the probability of either event alone. Other well-known biases introduced by the representative heuristic are the *gambler's fallacy* and *base-rate neglect*. The gambler's fallacy is the belief that when a series of trials all have the same outcome then soon an opposite outcome will follow. This belief originates from the idea that random sequences of outcomes seem more representative of a sample space. Base-rate neglect is neglecting the relative frequency with which an event occurs. This is again illustrated by an example from Tversky and Kahneman, where a group of participants is presented with the following description of a person whom they know stems from a population of 30 engineers and 70 lawyers:

Dick is a 30-year-old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

This description is entirely uninformative with respect to Dick's profession. However, when participants were asked to indicate the probability of Dick being an engineer, the participants gave a median probability estimate of 50%, whereas the correct answer would have been 30%. The participants ignored the base-rate and simply judged the description as equally representative of an engineer or a lawyer.

The *control* heuristic represents the tendency of people to act as if they can influence a situation over which they actually have no control. For example, people value a lottery ticket they selected themselves more highly than an arbitrary one given to them, even though the probability of winning a prize is the same for both tickets [3]. This *illusion of control* can cause overestimation of probabilities.

We have seen that the use of heuristics can introduce cognitive biases in probability assessments. The most prevalent biases are said to be overconfidence and base-rate neglect [5]. Overconfidence is especially a problem with extreme probabilities, that is probabilities close to 0% or 100%. People find extreme probabilities hard to assess; they are less likely to be overconfident about probability judgements that lie more in the centre of the 0% – 100% range [129].

5.1.2 Selection and motivation

Ideally, for probability elicitation, an expert should be selected who has the necessary domain knowledge and who is familiar with assessing probabilities. However, due to the nature of expertise (it is by definition a scarce commodity), there is often not a very large pool of experts to choose from. When eliciting probabilities for probabilistic networks, it is best to select an expert who has also been involved in building the structure of the network, to prevent errors due to the possible existence of different definitions for certain variables. It is also better to have more than one expert involved [18, 129], since different experts have different kinds of knowledge, all of which should be incorporated in the assessment. Assessments by more than one expert can be handled in two ways: collect the assessment of each expert and combine the assessments into a single one, or have the experts come to a consensus. The first approach has the mathematical advantage of enlarging the sample space, but assumes that nothing is gained from sharing knowledge and thought among the experts. With the second approach, group interaction problems, such as dominance of one expert over the other or pressure for conformity, can influence the assessment. Research on the subject of group assessment suggests that an optimal number of experts is around three [18].

Once the experts are selected, the elicitation task is introduced and its purpose is explained. The elicitation task will often be part of a larger process of step-wise refinement [26], where the experts are first asked to provide only initial assessments. With these assessments, a sensitivity analysis of the probabilistic network is performed revealing the most sensitive parts of the network; the most sensitive probabilities can then be refined, and so on. Refinement of the most sensitive probabilities is done by using additional information obtained from other sources than the experts involved, such as research reports or other experts. It has been observed that experts may feel that the assessments they are asked to provide are not subjective opinions, but numbers that can be checked in every-day practice [126]. They then have the uncomfortable feeling that the assessments they provide should be “correct” and this makes them less willing to cooperate. It is therefore important to convince the experts that their assessments need only be accurate in the sense that they should represent the knowledge and judgement of the expert: there are no right and wrong answers. Also, it may reassure the experts when it is explained to them that their initial assessments will be subjected to a sensitivity analysis and that they can thereupon refine their assessments.

Experts should also be informed about the biases discussed in the previous subsection; knowledge of their existence might help in counteracting them.

5.1.3 Training

Once an expert has been selected and is willing to cooperate, he has to learn the art of probability assessment. To this end, the expert should first of all become familiar with the concept of probability and should learn to express his knowledge in the format required by the elicitation method used. Part of the training is done with probabilities for events whose frequencies can be checked. This allows for exposing biases in the expert's assessments and to practice the elicitation method. Several elicitation methods and representation formats can be tried to see which best fit the task, the experience and preferences of the expert.

Feedback of the true frequencies of the events or which probabilities are assessed will help experts calibrate their responses [5], that is, will teach them to make assessments as close as possible to reality. However, care should be taken to not discourage the experts by the confrontation with their frequent mistakes. The events for which probabilities need to be assessed in a probabilistic network are often unobservable, making feedback impossible. The expert must, however, also become an expert at making probability judgements in this domain and part of the training should therefore be done with probabilities from the domain of the probabilistic network [39].

The amount of time spent on training depends on available time and other constraints. At the end of the training period, however, the expert should fully understand and feel comfortable with the methods to be used.

5.1.4 Structuring

Before the actual elicitation takes place, several issues need to be addressed. The definitions of the variables and values for which probabilities are to be assessed should be documented so that this information can be easily and promptly conveyed to the expert during the elicitation. For probabilistic networks this documentation will already be available from the construction of the graphical part of the network. Since probability elicitation is often done with the expert who was also involved in the construction of the graphical part, the expert will already be familiar with these definitions; it is, however, always a good idea to keep the documentation of definitions of variables and their values at hand during the elicitation interviews.

After the important variables and values are determined, the conditioning circumstances that influence a variable's uncertainty need to be determined. For probabilistic networks, these conditioning contexts follow directly from the structure of the network. For each probability to be assessed, a question describing this probability should be prepared. To suppress overconfidence and overestimation, questions should be prepared for assessment of an event's probability as well as for its complement(s).

In addition to the choice of elicitation method, the elicitor is faced with the choice of how to present the expert with the questions describing the probabilities that need to be assessed and what format to use for the expert's answers. Whatever representation is used to describe the probabilities to be assessed, the associated questions should be clear and structured in such a way that

there is no doubt about the variable a probability pertains to and the conditioning contexts. An attractive format should be prepared for the questions and, if possible, a graphical format for the answers. Experience shows that experts dislike writing numbers for subjective probabilities [20], since numbers suggest an accuracy that experts feel they cannot provide. The experts prefer to check scales, or place a ‘×’ in a box, etc. We will address the issue of presentation of both questions and answers in further detail in Section 5.2. The preparation of the questions and answering format may require a large amount of time on the elicitor’s part, but it is certainly time well-spent (see Chapter 7).

5.1.5 Elicitation and documentation

Various people will be present during the actual elicitation interviews. First of all, there will be one or more experts involved, interacting during elicitation [129]. There should be at least one, but preferably two, elicitors present during the elicitation, not in the least to show the experts that the task has sufficient priority for the expert to take it seriously. The elicitor has several tasks:

- He has to clarify the inevitable problems of the experts with the interpretation of questions, definitions of variables and values, etc.
- He has to record all information stated by the experts that cannot be expressed in the answering format, but may still be of use. For example, if an expert is allowed to express *trends* between conditioning contexts (see Chapter 7), such as “the conditional probabilities for this variable given this context are 10% higher than for that context”, it should be carefully recorded what is meant by this trend. Also, if the expert has overestimated the probabilities that pertain to a single conditional probability distribution, such that their sum exceeds 100%, possible information he has stated on the range within which the probabilities should lie, can serve to adjust them.
- It may turn out that certain conditioning contexts necessary to estimate the probabilities of certain variables are incomplete, or that certain contexts turn out to be unnecessary. For probabilistic networks, this indicates that changes have to be made to the structure of the network; it is important to carefully record this information.
- For some probability assessments, the elicitor may expect that certain biases are easily introduced; he should then once more make the experts aware of the biases.
- The elicitor should watch the clock: the elicitation is more taxing for the expert than for the elicitor and therefore sessions should not exceed one hour [20].

Despite the mentioned tasks, the elicitor should avoid coaching the expert and taking too much control [20, 129]; the expert should feel relaxed, not challenged, for he is the expert and the elicitor is not. The elicitation method that is used should be straightforward, easy to handle, and not difficult to learn [127]. The various elicitation methods commonly used will be discussed in some detail in Section 5.3.

5.1.6 Verification

When all required probabilities have been assessed, the elicitor should verify them. Verification is the process of checking whether the probabilities provided by the expert are well-calibrated (conform to observed frequencies), obey the laws of probability (are coherent) and are reliable [43]. Checking whether the assessments are conform “reality”, is often impossible, since the events for which the probabilities are assessed are often unobservable. Regarding coherence, we can check whether all probabilities that should sum to 100% indeed do so. It is convenient to do this check during the elicitation.

Test-retest reliability [39] tests whether the expert agrees with his own assessments, that is, whether the expert would provide the same estimates when asked for the same probabilities again. However, when dealing with probabilistic networks, the number of probabilities to be assessed is so large, that it is infeasible to assess them more than once. Instead of testing the reliability of separate assessments, entire probability distributions can be considered. As most probabilities are conditional, the expert can be shown the assessed probability distributions for a certain variable given different conditioning contexts and be asked to check whether the relationships for these different contexts are as he would expect. If not, the expert can adjust some of the assessments. Edwards [40] calls this an *antecedent conditions check* and he experienced that when his expert took these relationships into account during elicitation, the probabilities had a high test-retest reliability. We observed that our experts spontaneously mentioned these relationships, or trends, during elicitation (see Chapter 7).

An indication of the validity of the assessments can also be obtained by entering observations into the probabilistic network and computing the effect of the observations on the probabilities for certain variables of interest. The outcomes for these variables can then be checked against available data or presented to the expert.

5.2 Presentation

The presentation issues to be addressed for probability elicitation concern the representation format of the required probabilities, the description format of the questions to be asked and the answering format. Although we are interested in probabilities, the probability format is not necessarily required for the communication with the expert. The experts can, for example, be asked to provide odds or log-odds, or the most familiar competitor of numerical probability, verbal communication of uncertainty, can be used. When dealing with relatively probable events, probabilities or percentages may be intuitively convenient to experts, but in dealing with rare events, odds or log-odds may be easier because they avoid very small numbers.

Regardless of the format used for uncertainty, the required assessment can be described to the expert in different ways. The description format used should be conceptually simple and compatible with the expert’s abilities. When probabilities are chosen as the format for uncertainty representation, the required probabilities can be described, for example, in mathematical notation.

Example 5.1 Consider the domain of oesophageal carcinoma. We focus on the probabilities concerning the length of the tumour in the oesophagus of an arbitrary patient presented with

oesophageal carcinoma. In mathematical notation, the probability that an arbitrary patient with oesophageal carcinoma has a tumour longer than 10 *cm* would be presented as:

$$\Pr(\text{Length} > 10).$$

□

However, only experts who are very familiar with this notation will be able to completely understand it, especially when considering conditional probabilities where the meaning of what is represented on either side of the conditioning bar can be confusing.

Example 5.2 Again, consider the domain of oesophageal carcinoma. We now focus on the probabilities concerning the passage of food through the patient's oesophagus, which depends on the length of the carcinoma, its shape, and whether or not it is circular. In mathematical notation, the probability that an arbitrary patient with oesophageal carcinoma can swallow only liquid food, given that he has a polypoid, circular oesophageal carcinoma of more than 10 *cm* would be presented as:

$$\Pr(\text{Passage} = \text{liquid} \mid \text{Circ} = \text{circular} \wedge \text{Shape} = \text{polypoid} \wedge \text{Length} > 10).$$

□

People unfamiliar with the notation of conditional probability can easily get confused about the meaning of what is represented on either side of the vertical bar.

Another way of describing the required probability to an expert is to use the frequency format [46]. This format builds on the observation that registering occurrences of events is a fairly automatic cognitive process requiring little conscious effort. The basic idea is to transcribe probabilities in terms of frequencies, thereby converting abstract mathematics into simple manipulations on sets that are easy to recall and visualise.

Example 5.3 The probability presented in the example above using mathematical notation is described using the frequency format in the following way:

Imagine 100 patients with a circular, polypoid oesophageal carcinoma of more than 10 *cm*. How many of these patients will be able to swallow only liquid food?

□

Gigerenzer *et al.* argue that cognitive biases are merely artifacts of the presentation format and that the frequency format serves to suppress biases such as base-rate neglect, overconfidence, and the conjunction fallacy [46]. Overestimation of probabilities is reduced by assessing them as frequencies, because then people are more likely to be aware whether the sum of their assessments exceeds 100. The conjunction fallacy tends to disappear, because the frequency format appears to help people avoid choosing the most plausible description. For example, when asked “Out of 100 people like Linda, how many are bank tellers?” and “Out of 100 people like Linda, how many are bank tellers and active in the feminist movement?” (see also Subsection 5.1.1), most people correctly answered the latter with a smaller number.

Although the frequency format is easier for people to understand and apparently less liable to lead to mistakes, it is not always intuitively appealing. This is, for example, the case in

domains where experts find it impossible to imagine 100 occurrences of a rare event. The domain of oesophageal carcinoma is such a problem domain. Since oesophageal carcinoma is a low incidence disease in the Netherlands, the experts consulted often found it impossible to imagine 100 patients having the same characteristics.

Although the frequency method cannot always be applied, the idea of transcribing probabilities in words can be exploited in various other ways.

Example 5.4 The probability presented in the example above using the frequency format is transcribed without frequencies in the following way:

Consider a patient with a circular, polypoid oesophageal carcinoma of more than 10 *cm*. How likely is it that this patient will be able to swallow only liquid food? □

The final presentation issue concerns the format in which experts are required to give their answer. This format not only depends on the choice of uncertainty representation, but also on the choice of elicitation method. As we will see in the next section, some methods will require a verbal response, whereas others require an expert to, for example, mark a scale.

5.3 Methods

With the term *probability elicitation method*, we denote any aid that is used to acquire a probability from an expert. Generally, a distinction is made between *direct* and *indirect* methods. With direct methods, experts are asked to directly express their degree of belief as a number, be it a probability, a frequency or an odds ratio. For expressing probabilities, however, people find words more appealing than numbers. This is probably because the vagueness of words captures the uncertainty they feel about their probability assessment; the use of numerical probabilities can produce considerable discomfort and resistance among those not used to it [129]. Since, in addition, directly assessed numbers tend to be biased, various indirect elicitation methods have been developed. With these methods an expert is asked not for a direct assessment but for a decision from which his degree of belief is inferred; the use of an indirect method avoids having to explicitly mention probabilities for those who do not have clear intuitions about them [83]. For most methods, visual aids have been developed to make the elicitation easier on the experts.

In this section, we review the most commonly used methods for the elicitation of probabilities. These methods can be roughly divided into three categories:

- probability-scale methods;
- gamble-like methods;
- probability-wheel methods.

A probability-scale method is a direct method, where the expert is asked to indicate his degree of belief on a scale. The probability-wheel and gamble-like methods are indirect methods, since they require a decision instead of a number from the expert. We will devote a subsection to each of these categories and another subsection to some less known methods for probability elicitation we encountered in the literature.

5.3.1 Probability scales

A well-known direct method of elicitation is the use of a numerical probability scale such as the one shown in Figure 5.1. A probability scale can be a horizontal or vertical line with several anchors. In Figure 5.1, we have anchors denoting 0%, 25%, 50%, 75%, and 100% probability.

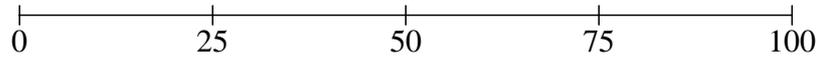


Figure 5.1: A numerical probability scale

For each probability that is to be assessed the expert is asked to mark the “correct” position on the scale. A separate scale is used for each probability. The indicated probability can be determined by measuring the distance between the mark and 0% on the scale. The expert should mark the scale in such a way that it is clear what position on the scale he is indicating, for example by using a small line or a carefully centred ‘×’, instead of circling the scale. The basic idea of the scale is to support experts in their assessment task by allowing them to think in terms of visual proportions rather than in precise numbers.

In addition to the horizontal probability scale, there exist also vertical scales and scales with a different number of anchors. Advantages of using a probability scale are that it is easy to understand and use and provides a fast method of elicitation, thereby allowing for elicitation of large numbers of probabilities. However, assessments made using a probability scale tend to be inaccurate and prone to scaling biases such as centering and spacing effects [129]. The *centering effect* describes the tendency of people to use the middle of the probability-scale; if people aesthetically divide their responses over the scale, this is termed *the spacing effect*. The spacing effect seems to originate from people’s tendency to organise perceptual information so as to optimise visual attractiveness. Note that the spacing effect cannot occur if a different scale is used for each separate assessment. Also note that the probability scales discussed are linear scales and therefore do not allow for elicitation of very large or very small probabilities. The use of a logarithmic scale would solve this problem. It should be kept in mind, however, that experts’ subjective scales are naturally equal-interval linear scales, not logarithmic scales [136].

5.3.2 Gamble-like methods

When people find it hard to express their degree of belief about some event as a number, their judgemental probability can be inferred from their behaviour in a controlled situation [5]. Indirect methods of probability elicitation such as, for example, the gamble-like methods are designed to represent such a controlled situation. The gamble-like methods for eliciting probabilities originate from the *Standard Gamble* introduced by Von Neumann and Morgenstern [128] as an indirect method for utility elicitation. The basic idea behind a gamble-like method is that the expert is presented with a choice between two lotteries. For one of the lotteries, the probability of winning corresponds to the probability of the event to be assessed; the probability of winning in the other lottery is set by the elicitor. The latter probability, or the associated price, is varied until the expert is indifferent about the two lotteries, whereupon the probability of the event to be assessed can be determined.

With a gamble-like method an expert is not required to give a probability assessment, but may instead compare a complicated concept with an event that does have meaning such as winning a lottery or a bet. We can distinguish two types of gamble. In the *certain-equivalent* gamble, a *sure thing*, that is, a 100% chance of winning, is compared to a lottery; the *lottery-equivalent* gamble consists of comparing two lotteries. From the choices made by the expert, the subjective probability for the associated event is inferred. Gamble-like methods can be presented to the expert graphically with the help of a decision tree depicting the possible alternatives, probabilities, and outcomes. The concept of decision trees, along with the symbols used, will have to be explained to the expert. When the expert fully understands the drawings, the elicitation process can proceed.

We will give an example of both variants of the gamble-like method. For each example we will explain what choices the expert has and how to determine the desired probability from his answers. In the first example we will also briefly explain the decision tree.

Example 5.5 Again consider the domain of oesophageal carcinoma. We focus on the probabilities to be elicited from the domain expert concerning the length of the tumour in the oesophagus of an arbitrary patient presented with oesophageal carcinoma. For ease of exposition we take the variable *Length* to be a binary variable with values $\leq 6\text{ cm}$ and $> 6\text{ cm}$. The probabilities required are the probability of a patient having a tumour with a length of 6 cm or less, and the complementary probability of the patient having a tumour longer than 6 cm.

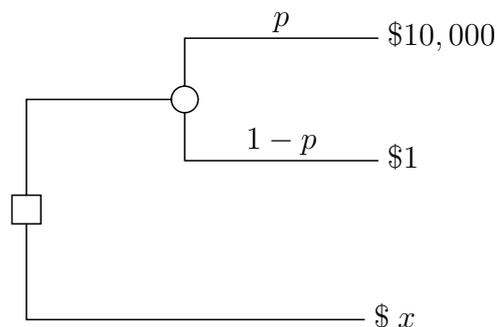


Figure 5.2: A certain-equivalent gamble

We will first consider a gamble with a certain equivalent, as depicted in Figure 5.2. Here the domain expert is presented with the following choice, indicated by a box (the decision node) in the figure:

- either enter a lottery where the pay-off (\$10,000, resp. \$1) depends on the “true” probability p of an arbitrary patient having a tumour of more than 6 cm,
- or accept a certain amount of money x set by the elicitor, instead.

The circle in the above figure indicates an uncertain event: with probability p the expert will earn \$10,000, and with a probability of $1 - p$ only \$1. The idea is that the elicitor varies the amount of money x in the certain equivalent until, for some value x' the expert is indifferent about the two alternative choices. In that case it is assumed that the expected value for both alternatives is

the same. We can then compute the probability p that the patient has a tumour of more than 6 cm from $x' = 10,000 \cdot p + 1 \cdot (1 - p)$. \square

A major drawback of this version of the gamble-like method is that elicited probabilities tend to be highly influenced by the *risk-attitude* of the expert. Some people are risk-seeking in the sense that they tend to choose a less probable alternative if it has a potentially more favourable outcome; other people tend to be risk-averse and will, for example, be more inclined to go for the certain outcome. Always going for the “sure thing” is known as the *certainty effect* [74].

A gamble with a lottery equivalent is less influenced by risk-attitudes. With this version of the gamble-like method, the expert is asked to choose between two lotteries; the price received upon winning (or losing) is equivalent for both lotteries.

Example 5.6 Consider again the example from the domain of oesophageal carcinoma, dealing with the elicitation of probabilities concerning the length of the tumour in an arbitrary patient with oesophageal carcinoma.

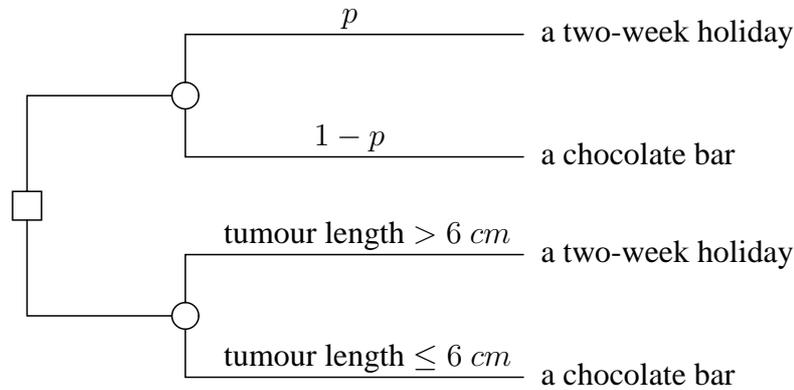


Figure 5.3: A lottery-equivalent gamble

When presented with a lottery equivalent gamble, the domain expert has the following choice:

- either enter the lottery where the outcome depends on some probability p set by the elicitor,
- or enter the lottery where the outcome depends on the probability of an arbitrary patient having a tumour of more than 6 cm.

In this lottery-equivalent gamble the probability p is varied until the expert is indifferent about the two alternatives. Again assuming that in that case the expected value of both alternatives is the same, we compute the probability p that the patient has a tumour of more than 6 cm from

$$\begin{aligned}
 & p \cdot \text{value}(\text{a two - week holiday}) + (1 - p) \cdot \text{value}(\text{a chocolate bar}) \\
 & = \\
 & \text{Pr}(\text{tumour length} > 6 \text{ cm}) \cdot \text{value}(\text{a two - week holiday}) + \\
 & \text{Pr}(\text{tumour length} \leq 6 \text{ cm}) \cdot \text{value}(\text{a chocolate bar})
 \end{aligned}$$

where *value* is a subjective measure of how valuable the outcome is to the expert. When the expert is indifferent, p directly represents the probability of a patient having a tumour longer than 6 cm, that is, $p = \text{Pr}(\text{tumour length} > 6 \text{ cm})$. \square

An advantage of this latter gamble over the former is that it directly presents the probability of interest and is less bothered by risk-attitudes. In addition, rewards can be expressed in terms other than money. As the gamble-like method does not require an expert to provide a probability assessment, it is considered to suppress some of the cognitive biases described in Subsection 5.1.1. However, the certain-equivalent gamble is easily influenced by risk-attitudes, which causes the probability derived from this method to be unequal to the expert's subjective probability, thus introducing a bias.

Gamble-like methods are not very expert-friendly methods. The methods are complicated to learn as they do not correlate with experts' usual cognitive processes. Also, experts may feel confronted with lotteries that are hard to conceive because of the rare and unethical situations they represent, like, for example, winning a two-week holiday if a patient dies [126]. Another drawback is that these methods are very time-consuming; they tend to take a lot of time per probability which makes them less suitable for assessing the thousands of probabilities required for probabilistic networks.

Studies that used the discussed elicitation methods for utility elicitation, report the consistent finding that numbers elicited with a probability scale are significantly lower than those elicited with the Standard Gamble [108, 115, 122]. Also, values obtained with the certain-equivalent gamble are consistently lower than for the lottery-equivalent gamble [74]. We are unaware of similar studies using the elicitation methods for probability elicitation.

5.3.3 Probability-wheel methods

An indirect method that is not influenced by risk-attitudes is the probability-wheel method. A probability wheel is a wheel-of-fortune-like wheel with two differently coloured sections. The sizes of these sections are adjustable and there is a pointer attached to the center of the wheel. An example of a probability wheel is shown in Figure 5.4.

Example 5.7 Using the same example as before, the expert is now asked which of the following events he considers most likely:

- the length of the tumour of an arbitrary patient with oesophageal carcinoma is more than 6 *cm*,
- *or*, after spinning the pointer, it will land in the red section.

The size of the red section of the probability wheel is adjusted by the elicitor until the expert considers the two events to have equal probability. The probability of an arbitrary patient having a tumour longer than 6 *cm* now equals the proportion of the probability wheel that is coloured red. □

The probability-wheel method has several drawbacks. The method tends to be very time-consuming, even infeasible when hundreds or thousands of probabilities are needed, as for probabilistic networks. Also, the method is quite close to direct estimation as the expert may recognise that the judgements he is asked to make are disguised assessments of the proportion of red showing on the wheel [129]; the advantage of suppressing judgemental biases, therefore,

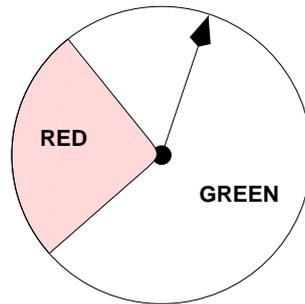


Figure 5.4: A probability wheel

may disappear. The method is not suitable for assessing very large or very small probabilities, for it will be difficult for an expert to distinguish between a very small red section and an even smaller red section. The advantage of probability wheels could be that they help experts visualise probabilities, but definitive conclusions from research on this are lacking [5].

5.3.4 Other methods

In this subsection we will briefly describe two other, very different and less-known methods for probability elicitation encountered in the literature. With the first method, experts are allowed to express their knowledge about uncertainties in any form they prefer and not necessarily in numbers. The second method requires experts to make pair-wise comparisons between events.

Druzdel and Van der Gaag [36] presented a method for probability elicitation where experts are allowed to provide both qualitative and quantitative information, whichever they are most comfortable with. The assumption underlying this method is that in the hyperspace of all possible probability distributions over the set of variables under consideration, one of these distributions is the “true” one. The information provided by the experts can be looked upon as a set of constraints used to diminish the hyperspace of possible distributions. These constraints are put in a canonical form resulting in a system of (in)equalities with constituent probabilities as unknowns. From these inequalities an upper and lower bound can be computed for any probability of interest. For the interval between upper and lower bound a second-order distribution is computed to determine the point within the interval that is most likely to be the actual probability. This second-order distribution is found by sampling from the distribution hyperspace and checking for each selected distribution whether it is a solution for the system of (in)equalities.

Another method, originally designed for utility elicitation, is the *analytical hierarchy process* [109]. With this method an expert is presented with all possible combinations of pairs of events for which utilities are to be assessed. When the method is used for probability elicitation, the expert is asked to compare, for each pair, the two events and to indicate the relative likelihood of events A and B using the scores shown in Table 5.1. This method has the advantage that experts are not required to explicitly state probabilities. Another advantage is that consistency of the expert’s statements can be easily checked, for the result from the comparisons should be a transitive ordering of events. However, using this method for probability elicitation for probabilistic networks poses two problems:

- The number of comparisons to be made exceeds, by far, the number of probabilities to be assessed. For example, the assessment of a mere 100 probabilities would require an expert to make $\binom{100}{2} = 4950$ pairwise comparisons of events.
- A lot of the events will differ so much that they are hard to compare for an expert.

Besides the problem of the great number of comparisons to be made, rather un insightful statistical methods are required to compute the probabilities from the results of the comparisons.

<i>score</i>	<i>relative likelihood</i>
1	<i>A and B are equally likely</i>
2	<i>undecided between 1 and 3</i>
3	<i>A is weakly more likely than B</i>
4	<i>undecided between 3 and 5</i>
5	<i>A is strongly more likely than B</i>
6	<i>undecided between 5 and 7</i>
7	<i>A is very strongly more likely than B</i>
8	<i>undecided between 8 and 9</i>
9	<i>A is absolutely more likely than B</i>

Table 5.1: The scale for pair-wise comparisons

5.4 Discussion

We discussed various issues that are to be taken into consideration when faced with the task of probability elicitation. We saw that probability judgements are prone to bias and that several elicitation methods have been developed to aid an expert in assessing probabilities, thereby suppressing, to some extent, these biases. It is clear that an important motivation for choosing a particular probability elicitation method is the ease with which both elicitor and expert understand and use the method. Moreover, the time an expert has available can limit the choice of methods. There will often be a trade-off between available time and the precision required, since the methods that are said to provide the most precise results are also the most time-consuming. Some people doubt however, that this trade-off really exists [63]: the use of gambles might not result in assessments that are as good as is believed, and faster methods such as the probability-scale methods might produce results that are better than believed.

While some of the phenomena reported in the heuristics and biases literature are real, reliable and reproducible, they may not be relevant, that is, they may not apply to the situation in which thousands of probabilities need to be assessed for a probabilistic network. For example, some biases, such as the conjunction fallacy, cannot arise during elicitation of probabilities for probabilistic networks [3]. Edwards [39] gives another three arguments why some of the phenomena may be irrelevant to probability elicitation for probabilistic networks. The first is domain expertise: for the elicitation of probabilities, experts are used who presumably know all there is to

know about the subject matter of the probabilities being judged. The studies concluding that humans are typically overconfident when providing probability estimates, arrive at that conclusion using general knowledge (almanac) questions and student participants that are often not trained in estimating probabilities. It is not at all clear that these results can be generalised to experts making assessments pertaining to their expert knowledge. Weather forecasters, for example, turn out to be very well calibrated indeed [38]. Another argument is probability judgement expertise: judging probabilities is something that can be learned. An expert who has done some training in estimating probabilities will find it easier to translate his knowledge and experience into probability judgements. The third reason is the possibility of consistency checks such as the sum checks and antecedent checks discussed in Subsection 5.1.6. These checks can be used during elicitation to provide the expert with information based on which he can, if necessary, reconsider his judgements.

When probability elicitation is seen as part of a stepwise refinement procedure, fast elicitation methods can be used to get initial rough estimates of the desired probabilities; sensitivity analysis methods [26] are then used to determine to which variables in the network the outcome is very sensitive. The focus of precise probability elicitation can then be put on the most sensitive parts of the network. Another important issue to keep in mind is that the networks are used to support a decision maker. They should at least improve the situation in which they are to be used, which means they do not always have to be 100% correct [39].

We are unaware of any systematic experimental evaluation studies of the different elicitation methods, especially in view of probabilistic networks; the results of the considerable number of empirical comparisons of methods do not show great consistency [83]. It is clear that a lot of research necessary to be able to decide on the best elicitation method, still has to be done. What is lacking are large multi-method studies where experts are asked to assess a large number of probabilities with every single method. It is important to get ecologically valid results, that is, results based on behaviour that is relevant to a real-world situation. Such results can provide for a meta-analysis about when to use which method and what methods not to use at all.

Designing a New Elicitation Method

In Chapter 5 we reviewed a number of probability elicitation methods that were designed to overcome, to at least some extent, the difficulties people experience when assessing probabilities. These methods, unfortunately, are often complicated and time-consuming. In this chapter we will present a new probability elicitation method tailored to fast assessment of rough initial estimates. The method is intended, in combination with sensitivity analysis procedures, to be part of a stepwise refinement procedure [26].

To be able to elicit a large number of probabilities in little time, probability assessment should be made easy on the experts. To this end, we look at their natural way of expressing probabilistic information. Except in situations where probabilities are objectively measurable, most people feel more at ease with verbal probability expressions than with numbers: when people communicate probabilistic information, they frequently do so in words rather than in numbers. In the assessment task, experts should therefore be allowed to communicate probabilistic information in words. Yet, it has often been argued that numbers are to be preferred over words. Words are more variably interpretable, the meaning being influenced by, among other things, context and personal opinions. In addition, verbal expressions are too vague, since different people translate the same verbal expression into different numerical expressions. Two assumptions underlie this argumentation against the use of words: the assumption that the correct way to interpret a verbal expression is by use of a numerical expression of probability and the assumption that numerical probability expressions are always interpreted in the same way. Uncertainty, however, is always dealt with within a context. This context can be either explicit, or people will implicitly think of one. Context not only influences the interpretation of verbal probability expressions, it also influences the interpretation of numerical probability expressions. The interpretation of both verbal and numerical probability expressions is based on actions and consequences related to the stated probabilities. For example, if ‘a low probability of infection’ and ‘a low probability of death’ are interpreted differently as a result of the consequences involved, then so will ‘a 23%

chance of infection' and 'a 23% chance of death'. In addition, numerical expressions often act as categorical descriptions such as '70 - 80%', or as anchors for ordinal descriptions, for example 'less than a half percent', rather than as real numbers [71]. For expressing uncertainty, numbers may therefore be just as vague as words.

From these considerations, we conclude that experts should be allowed to state their assessments in whichever mode, verbal or numerical, they prefer. However, as probabilistic networks require numbers, all verbal assessments need a numerical translation. To ensure that for each context this translation is agreed upon by the expert, we propose the use of a response scale similar to a probability scale but with both verbal and numerical anchors. To design this scale, we studied the relation between verbal and numerical probability expressions. As we are not the first to study this relation, we will start by reviewing other researchers' empirical studies on the use of probability expressions in Section 6.1. Results in favour of numbers as well as results in favour of words will be reported. Although no unequivocal conclusions can be drawn from these studies, they do provide us with several considerations, summarised in Section 6.2, to take into account when allowing the use of verbal probability expressions in communicating uncertain information. We took these considerations into account in our own studies, described in Section 6.3, in which we, unlike other researchers, never asked subjects to directly translate words into numbers or vice versa. From these studies a scale with both verbal and numerical anchors emerged, that constitutes the basis of the new elicitation method described in Section 6.4.

6.1 Modes of probability expression: previous studies

Many researches have studied the relationship between numerical and verbal expressions of uncertainty. In this section we review previous empirical studies on the use of probability expressions. Section 6.1.1 deals with the advantages of numbers over words for expressing probabilities. Subsequently, the advantages of words over numbers are discussed in Section 6.1.2.

6.1.1 Numbers versus words

Numbers have a persuasive advantage over words in the sense that they are precise, allow calculations, and have a fixed rank-order. Words are, in comparison, vaguer, do not allow calculations, and are more variably interpretable [131]. This disadvantage of words is apparent from the results of substantial empirical research studying numerical versus verbal probability expressions in general [12, 13], and studying the influence of context [8, 10, 84, 130] and severity of consequences [133] on the numerical interpretation of verbal probability expressions, more in particular. In these studies, student subjects, as well as experts were asked to translate numerical expressions into words and vice versa.

The empirical studies, for example, report the quite consistent finding of a great between-subject variability in the numerical values assigned to verbal probability expressions and a great overlap between the interpretations given for different words (cf. [8, 130]). Within-subject variability was however found to be small [12]. These results hold for both student and expert subjects, independent of whether or not a context was provided. Similar observations were found when subjects were asked to transcribe a graphical representation of a probability: much less

between-subject variability was found in the numerical probability expressions subjects used, than in the verbal expressions they gave. Within-subject variability in the use of expressions was again found to be small for both the numerical expressions and the verbal expressions [13].

Studies focusing on the influence of context on the interpretation of verbally expressed probabilities reported that the interpretation is indeed context-dependent, resulting in an even greater between-subject variability than when verbal expressions were presented without context. If winning a lottery is 'possible', entering the lottery may generally be considered a good decision to take, while if encountering a much disliked person at a party is 'possible', going to that party is generally not judged to be wise. Moreover, personal opinions about the consequences of the events referred to in the context presented result in individual variations in the meanings assigned to probability expressions. Some people may not mind meeting a disliked person or even enjoy the confrontation, others may definitely wish to avoid it.

Physicians are no exceptions to the above observations. When physicians were asked to translate verbal expressions into numerical expressions [11], they regularly gave different interpretations. When probabilities were communicated by verbal expressions, interpretations were also found to be highly variable, presumably because these were influenced by the severity of the consequences associated with the communicated information [81]. For example, 'low probability of infection' was interpreted differently than 'low risk of death'. Most of the authors of the studies referred to in this subsection conclude that physicians should use numerical, not verbal, expressions of probability (see also [85]): as verbal probability expressions may lead to confusion, numbers should be used.

6.1.2 Words versus numbers

Verbal expressions of probability are generally perceived as more natural than numerical probabilities, easier to understand and communicate, and better suited to convey the vagueness of opinions [131]. An interesting phenomenon, termed the 'communication mode preference paradox' was detected [41], however, when student and expert subjects in a study preferred to receive precise, that is numerical, information, yet preferred to express their own opinions in, vaguer, verbal terms. This preference was, however, not very strong, as subjects were willing and able to use both modes of expression (see also [94]). Other researchers found that one (student) subject out of three prefers numbers for both expressing and receiving information, saying that numbers are more precise, the second prefers words for both, and the third indeed exhibits the mentioned paradox [132].

Physicians rarely reason using numerical probabilities. As the use of numbers may wrongly suggest a precision of opinion [12], physicians prefer to use words in communicating probabilities to their patients [10, 71, 81]. In their communication, a variation of the preference paradox was found [10]: while physicians preferred to use words, thinking that their patients would understand words more easily, the patients preferred to receive information in numbers. Yet, when patients were given numerical information, they appeared to not understand the numbers as intended by the physicians. For example, a physician could state a 35% probability of having a disease and thereby intend to communicate a moderate probability; some patients might then understand that they had a considerable probability of indeed having the disease and be more

alarmed than the physician meant them to be, while others would understand it as less than fifty percent chance and overestimate their well-being (cf. [19]). Therefore, numbers should not self-evidently be preferred to words [10]. The two modes of communicating probability can both be used, as the argument that verbal expressions are too vague in meaning to be used in medicine is counter-balanced by indications that numbers have very little meaning for the average member of the public [86].

The between-subject variability found in the interpretation of verbal probability expressions varies from expression to expression. Expressions for the extremes of the probability range, that is: impossible and certain, are much less variably interpreted than expressions towards the middle of the range, such as possible or likely [66,67,84]. The use of qualifiers such as ‘very’ seems to introduce additional variability [116]. Comparison of a number of studies addressing the interpretation of verbal probability expressions led to the conclusion that, regardless of the population of subjects, the format of questions, instructions, and the context, for most verbal expressions the variation of mean assigned values was modest [84,86]. A study among medical subjects, moreover, concluded that although context influences the assignment of numbers to verbal probability expressions, it does not influence the rank order of the expressions [67]. In fact, an encouraging between-subject consistency was found in the rank ordering of verbal probability expressions by general practitioners [86] and individuals were found to have a relatively stable rank ordering of verbal probability phrases over time [12,67]. When subjects were asked to assign numerical interpretations to verbal expressions in a meaningfully ordered list, these assignments were less variable than their assignments to expressions in a randomly ordered list [52].

Taking into account the above experimental results, it seems justifiable to allow experts to continue to use verbal expressions if they prefer to, but enforcing more consistency of terminology [81]. Another argument that words are as suitable to express probabilities as numbers is found in the way information is processed. Whether people receive probabilistic information in verbal or in numerical form, appears not to influence the subsequent thought processes or actions based on the information. For example, when subjects were asked to provide answers for almanac questions and to indicate how confident they were of the correctness of their answers, the overall quality of these correctness forecasts in the two communication modes as well as the judgement processes were found to be similar [131]. The results for the communication modes did differ, however, in that with numbers, the 50% category was used much more often than the toss-up category was used in the verbal mode. Overconfidence was further found to be systematically more prominent with verbal than with numerical probability expressions. The overall conclusion seems to be that there are no grounds to prefer either numerical or verbal probability expressions as the better communication medium (cf. [13,41,48,87,117]).

6.2 Design considerations and goals

From the review in the previous section, we feel that there is sufficient justification for an elicitation method which allows the use of both numerical and verbal expressions of probability (cf. [60]). To prevent experts from using verbal expressions for which the variation in interpretation is found to be high, their choice of words should be limited to a pre-selected list. In designing such a list, we may take into account that differences in interpretation of verbal ex-

pressions may be reduced when the expressions are offered in a pre-defined rank order. Other studies addressing the relation between verbal and numerical probability expressions used lists of 18 expressions [13] or 19 [12, 52] or 34 in a long list and 14 in a shorter version [10], as few as two [48] or as many as 52 [84]. These lists were compiled by the researchers conducting the studies, which does not guarantee that people would actually use them. In fact, Zimmer [138] found that when subjects were asked directly for verbal descriptions of probability, the mean number of expressions used was 5.44. As it is easier to distinguish between a small number of expressions (“seven plus or minus two” [82]) than it is to demarcate the meanings of a long list of expressions [121], it is advisable to limit the number of verbal expressions used to a carefully selected list [12, 96]. For our elicitation method we will construct such a short list of rank ordered verbal expressions.

As the purpose of an elicitation method is the elicitation of *numerical* probability expressions, we have to find a way to translate the verbal expressions used by the experts into numbers. Other researchers have suggested the use of a table with translations between verbal expressions and their numerical meaning [44, 52]. We, however, do not want to fix the numerical interpretations of the verbal expressions to an explicitly specified number; we do want to provide a numerical translation for each verbal expression, but at the same time allow the expert to slightly adjust this numerical translation such that it better suites his opinion. In addition, we want to allow the expert to provide any numerical estimate, if he feels comfortable doing so. In Chapter 5, we saw that probability *scales* are used as an aid for probability elicitation. As these scales allow for fast elicitation, we feel that when probabilities are elicited from experts, the experts should be shown a scale, depicted graphically as a vertical line with numerical anchors on the one side and words on the other. When experts are more comfortable with numbers, they may refer to the numerical side of the scale and when they prefer to express their opinions in words, they may refer to the verbal anchors. The same scale may also be offered as reference for the interpretation of the output of a probabilistic network. Although probabilities elicited with a probability scale tend to be inaccurate (see Chapter 5), they are very useful as a first step in a step-wise refinement procedure.

In the following section we will describe the studies we undertook to arrive at a scale with both verbal and numerical anchors. In these studies, we took into account that the context in which expressions are elicited and presented may influence their interpretation. Unlike other researchers, in our studies we never asked subjects to directly translate words into numbers or vice versa, as we think that having to give such a translation is an artificial task, not true to actually performed cognitive processes (cf. [17]). Since our newly designed method was to be tested with medical experts, we included subjects with a medical background. The result of our studies is a set of expressions, whose numerical meaning is agreed upon and which together cover the whole range from zero to a hundred percent probability; this set of expressions is then used for the anchors on our response scale.

6.3 Our study

In this section we describe a series of four successive studies and discuss their results. The studies were set up to develop a set of verbal probability expressions, to be used in combination with

a numerical probability scale. The goal of the first study was to arrive at a list of commonly used verbal probability expressions. To this end, we asked subjects to generate such expressions. The second study was designed to check whether a stable rank order existed for the most commonly used expressions. We therefore asked (other) subjects to rank order the expressions from the first study. The third study was designed to determine the distances, or dissimilarities, between the expressions. This was done by asking subjects to make pairwise comparisons between each pair of expressions. The resulting distances were then used to determine how the words should be projected on a numerical probability scale. In the fourth study, we tested whether the translations that had resulted from the previous three studies for the verbal expressions were acceptable. To accomplish this, we did not ask subjects whether they thought that, for example, ‘a low probability’ equals ‘23%’; instead, we tested if, in a certain context, they interpreted ‘low probability’ the same as ‘23%’, that is, we tested whether people reacted the same to the verbal and the numerical expressions. In the fourth study, we therefore presented subjects with situations that required decisions; we then tested whether their decisions were influenced by the mode, verbal or numerical, in which probability information was presented. Examples of the questionnaires we used in the various studies are given in Appendix C.

6.3.1 The first study

In the first study we aimed at a list of commonly used verbal probability expressions. Most researchers use a dictionary or published articles to come up with a list of probability expressions. Since we had no reason to assume that such sources contain only the expressions most commonly used and, in fact, felt that these sources are more likely to list all linguistic possibilities, we designed a questionnaire.

Subjects There were 53 subjects in the study. Of these, 47 were students (Computer Science, Psychology, and Artificial Intelligence) and 6 faculty members; 23 were female and 30 male. The ages of the subjects ranged between 18 and 54 with a mean of 23 ($SD = 8.7$).

Procedure The subjects received the questionnaire. In the first paragraph they were asked for their cooperation in generating a list of commonly used verbal terms expressing (im)probability. Examples of the use of verbal probability expressions were given, such as “it is unlikely that I will pass my exam” or “I will probably go to Amsterdam this weekend”, to illustrate the basic idea. Instructions were given, in the second paragraph, to write down a list of terms judged suitable in situations where one wishes to express a degree of (im)probability. Subjects were reminded to only list expressions they thought were common, and to try them out for themselves in different virtual situations.

Results The 53 subjects together generated 144 different expressions. A mean of 8.2 expressions were given per subject ($SD = 4.1$). Of these 144 expressions, 108 (75%) were built from a probability term plus a modifier such as ‘very’ or ‘reasonably’. Some modifiers seemed synonymous, e.g. ‘almost possible’ and ‘nearly possible’, but we counted the phrases containing such modifiers as different phrases. Ninety-five expressions (66%) were used by only a single subject

and another 17 (11%) by only two subjects. Table 6.1 lists translations of the seven expressions that were used by 15 or more subjects (30%). The translations stem from the original Dutch phrases *mogelijk*, *waarschijnlijk*, *onwaarschijnlijk*, *zeker*, *onzeker*, *te verwachten*, and *onmogelijk*. The next often used term was given by 11 subjects, a couple of expressions were given by nine and eight subjects respectively, the remaining expressions had a frequency near one. Disregarding the modifiers in the composite expressions as a check of common use, that is, counting ‘almost possible’ and ‘nearly possible’ as ‘possible’ etc., the difference in frequency between the seven mentioned expressions and the remaining ones was even greater.

<i>expressions</i>	<i>frequency</i>
possible	38
probable	30
improbable	28
certain	25
uncertain	21
expected	18
impossible	15

Table 6.1: Verbal probability expressions and the number of times these were given by subjects in the first study ($n = 53$).

Discussion In studies into the use of verbal probability expressions, most other researchers used lists of expressions they compiled themselves by scanning literature or borrowing from others. Clark [17] proposed the method we used, that is, to ask subjects to generate lists of commonly used expressions. He had fewer subjects (20), who generated more expressions each, with a mean of 12.9, than our 53 subjects, with a mean of 8.2. The most frequently used expressions in his study were *certain*, *possible*, *likely*, *definite*, *probable*, *unlikely*, and *impossible*. So, he also found seven expressions, quite comparable to ours. He also found more ‘positive’ expressions, that is, from *fifty-fifty* towards *certain*, than ‘negative’ ones, that is, from *fifty-fifty* towards *impossible*.

In their attempt at translating verbal probability expressions into numbers, Mosteller and Youtz [84] advised to use the expressions *impossible* and *certain* for the two extremes, and *even chance* for the mid-point. To cover the rest of the range, they advised *probable* with modifiers. However, expressions with modifiers may give more rise to ambiguity than one-word expressions [116]. We therefore decided to use for our list *possible*, *probable* and *certain* plus their negations, and *expected* because that was used relatively often by our subjects.

We used the thus compiled list of seven frequently generated expressions for the next studies. As we wanted a term representing the center of the probability range, we added *undecided* (in Dutch *onbeslist*), to express *fifty-fifty* probability. This list of eight expressions neatly kept us within the advised range of seven plus or minus 2 [82]. We expected to resolve the asymmetry between the number of positive versus negative expressions in our next ranking and scaling studies.

6.3.2 The second study

The second study was set up to determine if a single, stable rank order existed for the eight verbal probability expressions found in the first study. In this study we only focused on a rank order. Distances between the expressions were established in the third study.

Design Subjects were asked to rank order the eight verbal probability expressions presented to them. We had a context and a no-context group. In the no-context group, the expressions were offered in isolation. In the context group, the probability expressions were embedded in a (Dutch) sentence describing a medical situation (for example: “It is certain that young people do not get varicose veins”).

Subjects Of the no-context groups, one group (group 1) consisted of 15 female and 11 male medical students. Their ages ranged from 19 to 45 years, with a mean of 21 (SD = 5). A second group (group 2) consisted of 19 female and 7 male social sciences students. Their ages ranged from 18 to 29, with a mean of 21 (SD = 2.5). Of the context groups, one group (group 3) consisted of 13 female and 8 male medical students whose ages ranged from 19 to 32, with a mean of 22.5 (SD = 5). The second group (group 4) consisted of 19 female and 3 male social sciences students; their ages ranged from 19 to 26, with a mean of 21 (SD = 1.6).

	<i>medical subjects</i>	<i>other subjects</i>
<i>no context</i>	group 1, $n = 26$	group 2, $n = 26$
<i>context</i>	group 3, $n = 21$	group 4, $n = 22$

Table 6.2: The subjects in the second study, with numbers of subjects per group.

Note that in both the context and the no-context groups, we had medical students and other (social sciences) students (see Table 6.2).

Procedure The subjects received a one-page questionnaire. At the top of the page, the task was introduced to them and instructions were given. The instructions were the same for both conditions, that is, the subjects were instructed to order the eight expressions, be they presented in isolation or embedded in a sentence, by assigning a ranking number to each of them. The number 1 was to be given to the expression denoting the highest level of probability and subsequent numbers to expressions denoting subsequently less probability. Assignment of the same ranking number to more than one expression was allowed (cf. [17]). Following these instructions, the eight expressions or sentences were presented, listed vertically. The presentation order was arbitrarily set to possible, impossible, uncertain, certain, probable, improbable, expected and undecided in the no-context condition and to probable, improbable, possible, undecided, impossible, uncertain, expected and certain in the context condition.

Data analysis A simple method of testing for differences in rank orderings provided by medical versus other subjects and for the context versus no-context groups, is to compute, for each group, the mean rank number assigned to a probability expression and to compare these means

by analysing variance (SD^2). For each of the eight verbal probability expressions, we analysed the between-group variance of the mean ranking numbers with a one-way ANOVA (ANalysis Of VAriance). Because we found significant differences in the means, we further analysed the data with PRINCALS (PRINCipal Components analysis by Alternating Least Squares), designed specifically for handling ordinal data.

	<i>no context</i> (<i>n</i> = 48)	<i>context</i> (<i>n</i> = 42)
certain	1.15 (0.94)	2.26 (2.46)
probable	2.70 (0.78)	3.29 (1.89)
expected	2.65 (0.85)	3.57 (1.49)
possible	3.81 (0.52)	4.56 (1.76)
undecided	5.69 (1.22)	5.40 (1.56)
uncertain	5.96 (0.87)	5.50 (1.46)
improbable	6.44 (0.84)	5.13 (2.16)
impossible	7.61 (0.99)	6.29 (2.33)

Table 6.3: Mean ranking numbers (and standard deviations) of the eight probability expressions, by the subjects in the no-context group (groups 1 and 2) and the subjects in the context group (groups 3 and 4).

For each dimension exposed in the data, PRINCALS can reveal a single ordering of the eight verbal expressions, that underlies the orderings provided by the different subjects. We assumed that all subjects gave their ordering along a single dimension, the level of probability. To test this assumption, we computed with PRINCALS a solution for two dimensions. If our assumption was correct, then the solution would have a high quality on only one dimension; on the other dimension the quality would be low enough to discard it. The quality of a solution, on a certain dimension, indicates how good the ordering computed by PRINCALS, for that dimension, represents the different orderings given by the subjects. For each dimension, PRINCALS indicates the quality of this ‘fit’ for every separate subject, as well as summarises the quality of a solution over all subjects. More detailed information on ANOVA and PRINCALS can be found in Appendix B.

ANOVA results The data of four subjects from group 1 and of one subject from group 4 had to be excluded from the ANOVA analyses, because the given ordering was incomplete. All rank orders were transformed to range from one to eight, where multiply used ranking numbers were encoded with their mean value, i.e. ‘1, 2, 2, 3, 3, 3, 4, 5’ was changed into ‘1, $2\frac{1}{2}$, $2\frac{1}{2}$, 5, 5, 5, 7, 8’, as the two 2s occupy positions 2 and 3, and the 3s occupy positions 4 through 6. The ANOVA analysis revealed that the four groups of subjects had assigned significantly different mean ranking numbers to five of the eight terms: possible, impossible, improbable, expected and certain. Post hoc tests using Tukey’s HSD-procedure showed that, for these five expressions, only the with-context and without-context group means differed significantly at $\alpha = .05$; there were no significant differences between medical subjects and other subjects for any of the expressions. We present the mean ranking numbers found in Table 6.3; since the only factor that contributed

to differences in means was context, the mean ranking numbers for the two no-context groups (groups 1 and 2), and for the two context groups (groups 3 and 4), are taken together.

Discussion From the ANOVA results, we concluded that context did indeed influence the rank order of the eight verbal expressions. The medical subjects and the others did not differ in the rank orders they produced. Our results are comparable to the results found in four related studies that did not include context (see Table 6.4).

	<i>no context</i> (this study) (<i>n</i> = 48)	<i>Tavana</i> (<i>n</i> = 30)	<i>Budescu</i> (<i>n</i> = 32)	<i>Clark</i> study 5.4 (<i>n</i> = 16)	study 5.2 (<i>n</i> = 16)
certain	1.15 (0.94)	1.05	–	–	1.36
probable	2.70 (0.78)	–	2.80	2.83	2.56
expected	2.65 (0.85)	–	–	2.62	–
possible	3.81 (0.52)	5.29	4.71	3.61	3.62
undecided	5.69 (1.22)	–	–	–	–
uncertain	5.96 (0.87)	–	4.94	5.38	5.94
improbable	6.44 (0.84)	–	6.22	6.45	6.83
impossible	7.61 (0.99)	8.00	–	7.87	7.90

Table 6.4: Mean ranking numbers (and standard deviations) of the eight probability expressions by the subjects in the no-context group (groups 1 and 2) and as reported by Tavana et al. [120], by Budescu & Wallsten [12] and by Clark [17], studies 5.4 and 5.2.

Considering that the data are ordinal, we feel uncomfortable with just the mean ranking numbers. For example, the high standard deviations for the mean ranking numbers given by the context group are difficult to explain by just looking at the means. A possible explanation of this phenomenon would be that the subjects did not rank the expressions on the single dimension of probability, but on another dimension as well. To check this, we perform an additional PRINCALS analysis.

PRINCALS results The data of none of the subjects had to be excluded from the PRINCALS analysis, as they had been from the ANOVA analyses, because PRINCALS can handle missing data. For both no-context groups, PRINCALS found a high-quality solution in one dimension (see Appendix B). Most subjects in these groups had provided rank orders along this one dimension; for two medical subjects and one other subject a second dimension was exposed. Because on inspection of their rank orders, there seemed to be no logical explanation for their order, we presumed that these three subjects had misunderstood the task and we excluded their data. The rank orders for the probability expressions given by the rest of the subjects in the groups 1 and 2 were quite the same.

For the two context groups, a solution in two dimensions was found. Upon inspection of the quality of the solution for the individual subjects, we found that 12 of the medical subjects scored high on the first dimension and had given comparable rank orderings; nine medical subjects

scored high on the second dimension. An additional PRINCALS analysis of the group of 12 subjects resulted in a one-dimensional high-quality solution. Inspection of the quality of the solution for the individual subjects from group 4 revealed that 11 of the 22 subjects had given comparable rank orderings and scored high on the first dimension; the other 11 subjects scored high on the second dimension. In an additional PRINCALS analysis of the 11 subjects scoring high on the first dimension, a one-dimensional high-quality solution was found.

On close examination of the rank orders given, the nine medical subjects and 11 others who scored high on the second dimension appeared to have judged the probability that the sentences in which the expressions were embedded were truthful statements, instead of judging the expressions themselves. As an illustration, one of these subjects judged **improbable** in the sentence “It is improbable that someone with tonsillitis does not have a sore throat” to express the highest level of probability and **possible** in the sentence “It is possible that someone faints from the heat” to express the lowest level of probability. Taking together the rank orders given by these apparently sentence-ranking subjects did not reveal an understandable pattern, however. We speculate that another factor had influenced these rank orders, possibly familiarity with the complaint for the medical subjects and everyday beliefs about such complaints for the other subjects.

Since the ANOVA analyses had shown that there were no differences between medical subjects and others, we did another PRINCALS analysis, taking these two groups together. For the resulting no-context group a high-quality solution was found in one dimension. The associated order of the verbal expressions is presented in the first column of Table 6.5. Taking the medical subjects and the others in the context groups together, and excluding the subjects who seemed not to have followed our instructions to rank order the expressions, we found a high quality one-dimensional solution; the expressions were ordered as shown in the second column of Table 6.5.

To conclude the study, we performed a final analysis over all four groups in the two conditions who had ranked the expressions on one dimension ($n = 72$). Their rank orders could indeed be summarised in one dimension. The associated order of expressions is shown in the third column of Table 6.5.

	<i>no-context group</i> (groups 1 and 2) ($n = 49$)	<i>context group</i> (groups 3 and 4) ($n = 23$)	<i>all subjects</i> ($n = 72$)
certain	1	1	1
probable	3	2.5	2
expected	3	2.5	3
possible	3	4	4
improbable	5.5	6	6
uncertain	5.5	6	6
undecided	8	6	6
impossible	7	8	8

Table 6.5: Rank order of the eight expressions of probability for the subjects in the no-context group (groups 1 and 2) and the subjects in the context group (groups 3 and 4), and all subjects together.

Discussion Surprisingly, the term **undecided** we had introduced to express fifty-fifty probability was ranked last by the subjects in the no-context group. The calculated means for this expression reveal a high standard deviation. Clearly, the interpretation of the expression **undecided** is not unambiguous. An explanation may be the order in which the expressions were presented to the subjects: **undecided** was the last expression in the list.

Our PRINCALS analysis showed that not all subjects in the context group rank-ordered the expressions: almost half of them gave rank orderings on a second dimension. This phenomenon was not found for the subjects in the no-context group. We therefore conclude that context does not highly influence the rank ordering of the expressions themselves, but context does seem to distract subjects from the actual task.

For ordinal data, the order of numbers is far more essential than the distances between the numbers. Computing means for ordinal data can, therefore, give a distorted impression of the data: atypical ranking numbers provided for an expression may affect the mean ranking number more than it affects the PRINCALS solution. Therefore, in our opinion, the PRINCALS analysis is more appropriate than the analysis of mean rank orderings. To conclude, we summarise the results of our second study as revealing the following rank order of our eight verbal probability expressions: **certain** and **impossible** at the extremes, with **probable**, **expected** and **possible**, in that order, expressing less probability from the certain-side down, and **uncertain**, **improbable** and **undecided** toward the impossible-side. This rank order is stable over all groups and independent of context.

6.3.3 The third study

As we were constructing a scale to be used in a probability elicitation method, and therefore required numerical translations of our verbal expressions, an order of the expressions alone was not sufficient. We had to establish the ‘distances’ between the expressions, that is, we had to establish whether two (or more) expressions were taken to mean almost the same or were quite distinguishable in meaning. To this end, we set up a third study, in which we asked subjects to rate differences between expressions. We expected to find that **certain** and **impossible** would be judged as extremely different, while **uncertain**, **improbable** and **undecided** might be judged as rather similar.

Subjects In the study we had two groups of subjects, again one group with a medical background (group 1) and one comparable group with another background (group 2). The subjects in group 1 were 28 students from the Department of Medical Biology. 12 of them were female and 16 were male. Their ages ranged from 19 to 25, with a mean of 20 ($SD = 1.5$). The subjects in group 2 were 56 Computer Science students. 13 of them were female and 43 were male; their ages ranged from 20 to 53, with a mean of 24 ($SD = 4$).

Procedure In the study we asked subjects for pairwise comparison, that is, for similarity judgements for all pairs of verbal probability expressions. For the eight expressions, there were 28 pairs. A similarity judgement was made by putting a mark for each pair of expressions on a 10 cm line, using for the extremes the expressions ‘exactly the same’ and ‘completely different’

as anchors. Each judgement was made on a separate sheet of paper. The order of presentation of pairs was random across subjects and, for a pair AB, half the subjects received the pair with A listed first, while the other half received B first. Subjects performed four trial runs before commencing with the real judgements.

Data analysis The judgement of (dis)similarity for each pair of expressions from each subject was scored in millimetres, read from a ruler placed against the 10 cm line. For each subject an 8×8 data matrix was constructed, in lower triangular form with zeros on the diagonals. The matrices were analysed with ALSCAL (Alternating Least-Squares SCALing). ALSCAL takes a matrix of distances between objects and computes the positions (coordinates) of the objects in some n -dimensional space, using Euclidean distance. ALSCAL can also compute a single solution for all matrices together. Since all probability expressions seemed to be comparable, we did an analysis in only one dimension. ALSCAL produces a list of x -coordinates for the eight probability expressions. These coordinates are computed such that their fit with the distances between the expressions given by the different subjects is as good as possible. More detailed explanation of ALSCAL can be found in Appendix B.

We mapped the coordinates of the expressions yielded by ALSCAL onto a scale, with as anchors the expressions (**certain** and **impossible**) representing 100% and 0% probability, respectively, and then using a linear transformation to calculate the probabilities of the other expressions (compare [120]).

ALSCAL results The initial ALSCAL analysis of the matrices of the medical subjects from group 1 showed that the matrices of two subjects deviated considerably from the calculated coordinates. Upon inspection of the matrices of these two subjects, the deviation seemed to be the result of a judgement of **certain** and **impossible** as very similar (a distance of 1 mm on the 10 cm line, where one would expect the full 10 cm). We removed the data from these two subjects and performed another analysis with the remaining 26 matrices. The coordinates of the eight expressions, returned by ALSCAL, are given in the left half of the leftmost double column of Table 6.6; the right half of this double column presents a linear mapping of these coordinates onto a probability scale from one to zero. The initial analysis of the matrices of the subjects from group 2 showed that the matrices of four subjects had to be removed because of their poor fit. The analysis with the remaining 52 matrices resulted in the coordinates of the eight expressions as given in the left half of the middle double column of Table 6.6. The right half of this column again presents a linear mapping of these coordinates onto a probability scale. We performed a final analysis over the 78 matrices from the two groups together. The right double column of Table 6.6 presents the results of this analysis, again with calculated probabilities.

Discussion The probabilities calculated for the expressions **probable** and **possible** are close to one another in the final analysis. Moreover, these probabilities are different, and inverted, for the two groups. We concluded from these observations that they could be taken as referring to the same range on the scale and that one of them can be removed. We left out **possible**, because its interpretation differs most between the two groups. The probabilities calculated for **undecided** are very different for the two groups as well. Since the expression was again (cf. study 2) not

	<i>group 1</i> (<i>n</i> = 26)		<i>group 2</i> (<i>n</i> = 52)		<i>all subjects</i> (<i>n</i> = 78)	
	coord.	prob.	coord.	prob.	coord.	prob.
certain	1.1950	1.00	1.2952	1.00	1.2738	1.00
possible	1.0897	0.96	0.8284	0.84	0.9105	0.86
probable	0.8409	0.87	0.9252	0.87	0.9043	0.86
expected	0.7239	0.82	0.7211	0.80	0.7133	0.79
undecided	-0.5972	0.32	-0.3730	0.41	-0.4394	0.38
uncertain	-0.7210	0.28	-0.8139	0.26	-0.7939	0.25
improbable	-1.0741	0.14	-1.0435	0.17	-1.0610	0.16
impossible	-1.4572	0.00	-1.5394	0.00	-1.5075	0.00

Table 6.6: Coordinates and calculated point probabilities for the eight expressions as given by the subjects from group 1, from group 2, and by all subjects together.

interpreted as intended, that is, for the mid-point of the probability scale, we decided to leave out this term as well. As this left us with a scale without a mid-point, we decided to add one term that can hardly be misunderstood, *fifty-fifty*, although it is arguable whether this is a verbal probability expression.

Upon close inspection of the matrices, we noticed that the positive-negative pairs *certain-uncertain* and *possible-impossible* were judged by most subjects as 100% dissimilar. Taking all expressions into consideration, *uncertain* and *possible* may be expected to be at some distance from the extremes *impossible* and *certain*. Our method of eliciting pair-wise dissimilarity judgements, however, seemed to have had artificially forced interpretation of the expressions towards the endpoints of the scale. We therefore feel justified to slightly reinterpret the calculated probabilities towards the mid-point, resulting in the list of seven verbal expressions with their point probabilities, presented in Table 6.7.

<i>verbal expression</i>	<i>probability</i>
certain	100%
probable	85%
expected	75%
fifty-fifty	50%
uncertain	25%
improbable	15%
impossible	0%

Table 6.7: The final seven verbal probability expressions plus their calculated probabilities.

6.3.4 The fourth study

Our fourth study was designed to test the translations of the verbal probability expressions into the numerical probabilities as established in the third study. To this end, we compared the de-

cisions subjects made when offered probability information in verbal form, to their decisions when presented with the information in numerical form. If the calculated point probabilities indeed have the same meaning as the verbal probability expressions, subjects will make similar decisions regardless of the mode of presentation. Not only the decisions subjects make will give an indication of how similar they think the verbal and numerical expressions are, but also their confidence in the correctness of their decision. For example, we expect that when subjects decide, with high confidence, to cancel an appointment when they are informed that rail-workers will ‘probably’ continue their strike, they also decide, highly confidently, to cancel their appointment when rail-workers continue their strike with 85% probability.

Subjects In the study, we had 123 subjects, being students and faculty members of the departments of Computer Science, Psychology, Artificial Intelligence, and Medicine. Of these subjects, 59 were female and 64 were male. The ages of the subjects ranged from 18 to 61, with a mean of 28 ($SD = 9.7$).

Procedure The subjects received a two-page questionnaire, with an introduction to the task they had to perform, followed by six decision situations. Each decision situation was described in two or three lines, such as:

Ms. T. has a non-serious physical complaint, which does however need treatment.
The probability that Ms. T. is allergic to the usually prescribed drug H. is about
Alternative drugs for her complaint are available, but these are less effective. Do you prescribe drug H. ?

Each of the six situations was followed by a table that contained either the seven verbal probability expressions or the seven numerical probabilities from Table 6.7. Each of the seven items was followed by “decision: yes/no” (to be circled) and by a 2-cm line on which subjects were to mark their measure of confidence in their decision (from complete to none). The subjects were instructed to mentally write, on the dots in the description, each of the expressions in turn, to make a yes/no decision for that hypothetical situation and to mark their confidence. Each subject thus made seven decisions plus confidence marks for each of the six situations.

We had four versions of the questionnaire. Version one started with three decision situations A, B and C, with verbal expressions, followed by three situations D, E and F, with the list of numerical probabilities. Version two contained the same six situations in the same order, but now with situations A, B and C with numerical probabilities and situations D, E and F with verbal expressions. In versions three and four the six situations were given in the order D, E, F followed by A, B, C, with version three starting with verbal expressions and version four with numerical probabilities. The tables of expressions and probabilities, each of which occurred three times in a questionnaire, were first given in the order we had determined in the third study, then twice in a different random order. A description of the six decision situations can be found in Appendix C.

Data analysis For each of the six decision situations we had a verbal and a numerical answering mode. In each mode, a yes or no decision was made for seven expressions. With these three

variables, that is, Mode, Decision and Expression, we constructed a $2 \times 2 \times 7$ matrix for each situation. The fields of the matrices specify the total number of subjects who made a certain decision in a certain mode for a certain category. For example: 34 subjects decided ‘no’ in the verbal mode for ‘fifty-fifty’. Similar matrices were constructed for the subjects’ confidence. We measured the confidence subjects had in their decisions by scoring their marks on the confidence line. Complete confidence was scored as 1.0, no confidence as 0.0. We computed both the sum of the confidences of all subjects together and the mean confidence for each mode-decision-category combination. This resulted in another two sets of six matrices with in each field the total and the mean confidence, respectively. We analysed our matrices using HILOGLINEAR, a log-linear analysis method. A log-linear model describes the (in)dependences between variables. For example, the model $\text{Expression} \times \text{Decision} \times \text{Mode}$ describes a possible three-way dependence between all three variables; the model $\text{Expression} \times \text{Decision} + \text{Mode}$ indicates that only Expression and Decision may be dependent, but that Mode is certainly unrelated to them. More detailed explanation of a log-linear analysis can be found in Appendix B.

Results We performed separate analyses for the decisions and for the confidence. We had to remove the data from 12 of the 123 subjects, because these subjects had misunderstood the assignment and had made a decision for only one expression in each situation. Of the 110 subjects left, 52 answered version 1 or version 3 of the questionnaire and 58 subjects answered version 2 or version 4.

For five of the six decision situations, the best log-linear model fitting the data was $\text{Expression} \times \text{Decision} + \text{Mode}$. In other words, the decisions were related to the probability expression used, but not related to the mode, verbal or numerical, in which the expression was given. For the sixth situation, the only model that fit the data was $\text{Expression} \times \text{Decision} \times \text{Mode}$. Knowing that variables are independent is more valuable than knowing that variables are *possibly* dependent. A model making no independence assumptions, such as $\text{Expression} \times \text{Decision} \times \text{Mode}$, contains no valuable knowledge and will therefore always fit the data. To account for small field frequencies, we performed a continuity correction on our data. With the corrected data, the same model as for the other five situations resulted, but it was not convincing. Upon close examination of the fields, we traced the sub-optimality to a single expression: the proportion of ‘yes’ to ‘no’ decisions differed by a factor 4 between the numerical expression 25% and the verbal expression uncertain.

For the matrices with the total confidence per cell, again, the model $\text{Expression} \times \text{Decision} + \text{Mode}$ was the best fitting model for four of the six decision situations. In other words, the confidence subjects had in their decisions, was related to the probability expression, but not related to the mode in which the expression was given. For one of the two remaining situations, the model $\text{Expression} \times \text{Decision} + \text{Mode} \times \text{Decision}$ was significantly better. However, after a continuity correction on our data the same $\text{Expression} \times \text{Decision} + \text{Mode}$ model was found to be the best model. For the other situation, the only fitting model was $\text{Expression} \times \text{Decision} \times \text{Mode}$. Performing a continuity correction did not result in another model. In this situation, we again found a large difference in the proportion of ‘yes’ to ‘no’ decisions between the two modes of one expression, 25% and uncertain. Deletion of this expression and a continuity correction again resulted in $\text{Expression} \times \text{Decision} + \text{Mode}$ for the best model.

Analysis of the matrices with mean confidence per cell did not provide us with very strong results; as field frequencies were so close together, almost any model would fit the data. From close examination of the matrices, we conclude that for the six decision situations there seemed to be no difference in the subjects' mean confidence in their decisions and no difference between the subjects' mean confidence for decisions in the verbal mode and in the numerical mode. Subjects seemed consistent in their confidence judgements over situations as well as in the two modes.

Discussion Our analysis showed that the decisions subjects made, for the six situations presented to them, depended only on the expressions used in the description of a situation and that the decisions were not influenced by the mode in which the expressions were presented. We did experience that one expression, 25% or *uncertain*, was causing problems with finding a good fitting model for some decision situations. This could indicate that to some subjects *uncertain* and '25%' have different interpretations and are not considered to be two different modes for the same expression. Indeed, to some people *uncertain* may mean anything less than a 100% certain, others could interpret *uncertain* as a 50% certainty. However, this problem only occurred in the situations where the expressions were presented in random order: it did not occur when the expressions were presented in order. In fact, the best fit was found for the best fitting models for the situations in which the ordered lists were presented. We conclude that when the probability expressions are presented in an ordered list, they will be interpreted as intended.

From the results of our analysis, we conclude that context influences the decisions people make. However, because we only found differences in decisions between decision situations and not per situation between the verbal or numerical mode, we conclude that the two modes are interchangeable. Our results therefore suggest that the agreement between the calculated point probabilities and the verbal probability expressions, as given in Table 6.7, is reliable.

6.3.5 Overall summary of results

The first three studies we performed, resulted in an ordered list of seven commonly used verbal probability expressions, which together span the whole probability range. Numerical equivalents for the verbal probability expressions were computed using the dissimilarities between expressions provided by subjects in the third study. Our studies differed from others in an important aspect: we did not ask people to translate numerical expressions into numbers or vice versa. In our opinion, asking for such a translation forces subjects to use two different mental representations of probability at the same time and to look for a mapping between the two. We addressed only one representation, thereby avoiding possible confusion. The fourth study was designed to test the validity of the computed translations. The finding that subjects made the same decisions, with the same confidence, irrespective of communication mode, justifies the tentative conclusion that this translation is acceptable.

There are some shortcomings to our study. The subjects used were Dutch and consequently we used Dutch words, which we translated into English for this thesis. Although we made no choices for the English words because dictionaries only give the one translation for each term, we cannot be sure that the connotations of the Dutch and the English words are similar. A replication with native English speaking subjects could verify this point. We are not satisfied with the way

the expression for the midpoint of the probability range was determined. Because our subjects did not write down such a term, we first introduced *undecided*. Its literal meaning may be fifty-fifty, but our subjects did not appear to interpret it this way. We then replaced it by *fifty-fifty*, which seems cheating on what may count as verbal. Moreover, because we introduced this term later, the assumed distance to the other terms was not established as it had been for the other terms.

6.4 The new elicitation method

The studies in the previous section were conducted with the intention of finding a small set of verbal probability expression, with numerical translations, that could be used as anchors on a response scale. This response scale would then be used as part of a probability elicitation method. In this section, we present the resulting response scale and discuss how we incorporated the various presentation issues, reviewed in Section 5.2, into our method. The use of the method in an actual elicitation process will be commented upon in Chapter 7.

Our method for probability elicitation from domain experts combines various ideas. Although several of these ideas were presented before by others, we combined and enhanced them to yield a novel and, as we will argue in the next chapter, effective elicitation method. The most important ingredients of our method are the response scale we present to domain experts, resulting from the studies in the previous section, and the presentation format of the event for which a probability needs to be assessed.

Recall from Section 5.2, that there are three presentation issues to consider in a elicitation process: the representation format of the probabilities, the description format of the questions presented to the experts, and the answering format. As representation format of the probabilities we use likelihoods. Using likelihoods rather than the often suggested frequencies helps to forestall difficulties with the assessment of a conditional probability for which the conditioning context is quite rare. The questions regarding the required probabilities are presented to the experts as fragments of text transcribing the probability to be currently assessed. Using a fragment of text to denote a probability circumvents the need to use mathematical notation. Especially for experts who are less familiar with the mathematical notation as in

$$\Pr(Invasion = T2 \mid Shape = polypoid, Length < 5 \text{ cm}),$$

we propose transcribing the requested probability with a fragment of text that should be understandable to them:

Consider a patient with a *polypoid* oesophageal carcinoma; the carcinoma has a length of *less than 5 cm*. How likely is it that this carcinoma invades the *muscularis propria* (*T2*) of the patient's oesophageal wall, but not beyond?

From this fragment it can be seen that experts are required to make a statement about likelihood. As we want to allow experts to make such likelihood statements using either verbal or numerical probability expressions, we present them with the expressions resulting from our studies in the previous section. We had chosen the probability scale as a fast method of elicitation; the expressions from Table 6.7 are put as anchors on a vertical probability scale. Experts are now allowed

to mark their assessment right next to a word or at any point between two words on the scale. Since the verbal probability expressions were explicitly intended as independent anchors on the scale rather than as fixed translations for the numerical probabilities, we decided to position the verbal probability expressions close by rather than simply beside the numerical anchors. We further decided to add the moderator “(almost)” to the most extreme verbal expressions to indicate the positions of very small and very large probabilities. The resulting response scale is the scale shown in Figure 6.1.

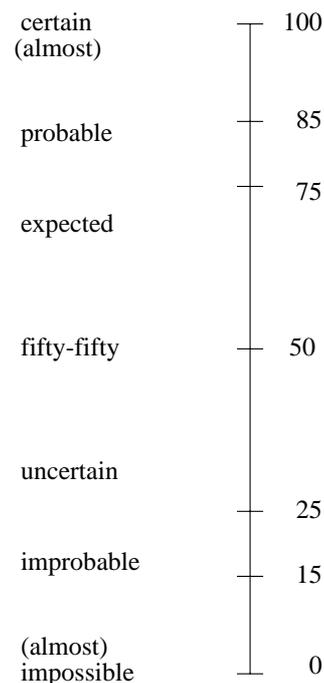


Figure 6.1: The response scale with both verbal and numerical anchors as used with our new method.

Having discussed the process of assessing a single probability, we will now discuss how we accommodate the assessment of a number of probabilities. To this end, domain experts are presented with a separate figure for every probability that needs to be assessed. On the right of the figure, our response scale is depicted; to the left of the scale is a fragment of text transcribing the probability to be assessed. By using a separate response scale for each assessment we can prevent the centering effect mentioned in Section 5.1.1. The figures pertaining to the various probabilities are grouped in such a way that the probabilities from the same conditional distribution can be taken into consideration simultaneously; they are presented in groups of two or three, if necessary on consecutive single-sided sheets of paper. An example is shown in Figure 6.2. Explicitly grouping related probabilities has the advantage of reducing the number of times a mental switch of conditioning context is required of the domain experts during the elicitation. It also allows the experts to immediately check the coherence of their judgements, for the assessed probabilities have to sum up to one, thereby possibly reducing overconfidence (see Section 5.1.1).

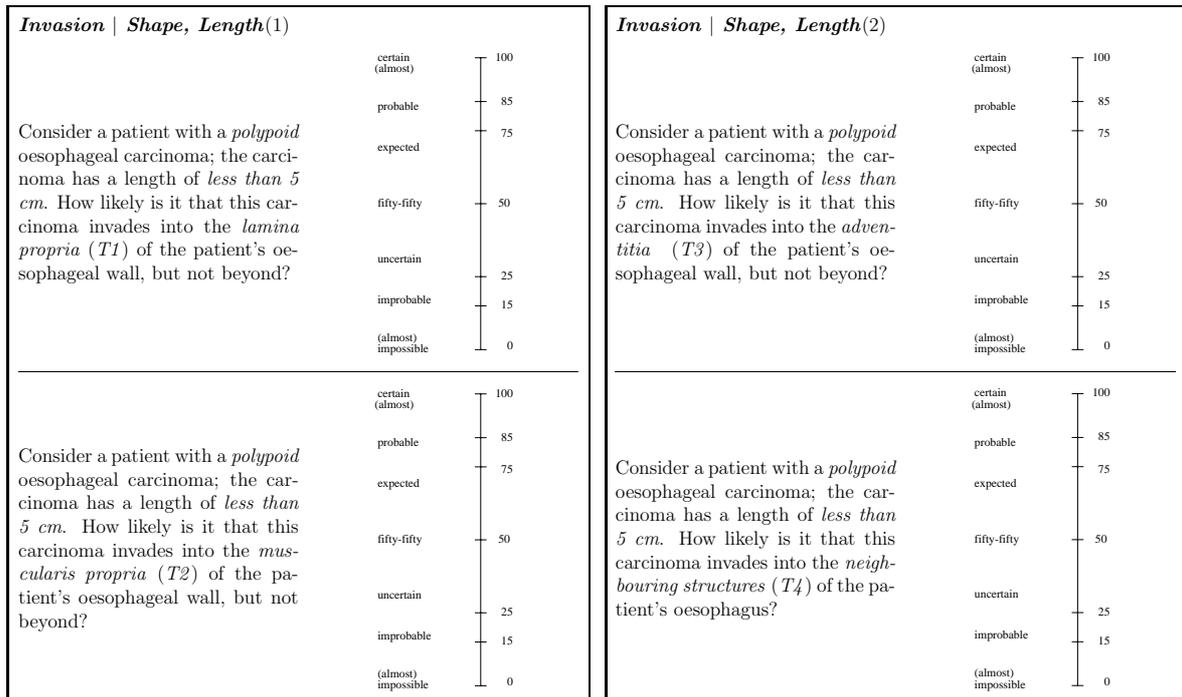


Figure 6.2: Two pages with all figures pertaining to the probability distribution for *Invasion* given a *polypoid* carcinoma with a length of *less than 5 cm*.

6.5 Conclusions

In this chapter, we discussed the design of a new elicitation method tailored to the fast elicitation of probabilities. With our new method, we present domain experts with a response scale with both verbal and numerical probability expressions as anchors. This allows experts to choose the representation mode that best matches their normal way of thinking in the situations described. We expect that during elicitation experts will prefer verbal probability expressions when they are more uncertain and that they will use numerical probability expressions when they feel safe doing so. By presenting (conditional) probability questions in text format, grouped together in such a way that experts are facilitated to make coherent judgements, and by presenting a verbal-numerical response scale for each response, we took as much care as we could to ensure the ecological validity of the task and the response mode. Experiences with the method in an actual elicitation process are discussed in the next chapter. Still, a more systematic study into the benefits of the use of the new method is called for.

Besides using our response scale as part of an elicitation method, we also envision it as a useful tool in the explanation of a system's output and reasoning-process.

Experiences with our Elicitation Method

To evaluate the quality of a newly designed method, it must be established what the purpose of the method is and to what extent this purpose is met in practice. The purpose of a probability elicitation method is to extract an expert's subjective beliefs. A probability elicitation method is a good elicitation method if the assessments provided by an expert indeed correspond to his subjective beliefs. Ideally, experts are well-calibrated and their subjective belief in a certain event corresponds to the frequency with which that event occurs in the physical world (whether or not this can be measured). In an ideal situation, a good probability elicitation method will therefore elicit well-calibrated assessments. When quantifying a probabilistic network, we consider the calibration of assessments to be of more importance than their correspondence to the expert's subjective beliefs. The purpose of *using* an elicitation method therefore conflicts with the purpose of the method itself. Therefore, to evaluate the quality of our new elicitation method described in Chapter 6, we will consider whether or not the purpose of the method's use is met, that is, we will evaluate the quality of the obtained probabilities.

Our elicitation method has been used with two domain experts from the Netherlands Cancer Institute, Antoni van Leeuwenhoekhuis, to quantify the network on oesophageal carcinoma described in Appendix A. Thus far, we have focused our elicitation efforts on the diagnostic part of the network which constitutes a coherent and self-supporting probabilistic network and currently includes 42 nodes with a total of 932 probability assessments. To assess the quality of these probabilities obtained with our new elicitation method, we conducted an evaluation study of the oesophagus network, using data, from the Antoni van Leeuwenhoekhuis, from 185 patients diagnosed with oesophageal carcinoma. The evaluation study focused on the diagnostic part of the network that provides for establishing the *stage* of a patient's carcinoma. This stage summarises the carcinoma's characteristics, its depth of invasion, and the extent of its metastasis, and is indicative of the likely outcome of treatment. We would like to note that the characteristics, depth of invasion, and extent of metastasis themselves are of interest rather than the stage derived

from them; focusing on the summarising stage, however, serves for gaining overall insight in the diagnostic part of the network.

In this chapter, we will discuss our overall experiences with the new elicitation method. To compare the use of our method with that of existing methods, we describe our initial experiences with probability elicitation for the oesophagus network in Section 7.1. In Section 7.2 we evaluate the use of our method in the construction of the oesophagus network; more specifically, we comment on the observations made by our domain experts. In Section 7.3 we present the results of the evaluation study of the network. Some concluding observations are given in Section 7.4.

7.1 Initial experiences with probability elicitation

In the construction of the oesophagus network, probability elicitation soon proved to be a major obstacle. As in many problem domains, numerous sources of probabilistic information seemed to be readily available. We collected data from historical patient records and we performed a literature review. Unfortunately, the Netherlands being a low-incidence country for oesophageal carcinoma, we were not able to compose an up-to-date, large and rich enough data collection to allow for reliable assessment of all probabilities required; after due consideration, we decided to save the collected data for evaluation purposes. Literature review also did not result in ready-made assessments. Although the literature provided abundant probabilistic information, it seldom turned out to be directly amenable to encoding in our network. Research papers, for example, often reported conditional probabilities of the presence of symptoms given a cause, but not always the probabilities of these symptoms occurring in the absence of the cause. Also, conditional probabilities were often given in a direction opposite to the direction required. For example, the statement “70% of the patients with oesophageal cancer are smokers” specifies the probability of a patient being a smoker given that he or she is suffering from oesophageal cancer, while for the network the probability of oesophageal cancer developing in a smoker was required. Another problem was that the characteristics of the population from which the information was derived were not properly specified or deviated seriously from the characteristics of the population for which the oesophagus network is being developed. Because of these and similar problems, hardly any results reported in the literature turned out to be usable for our network. The knowledge and personal clinical experience of the two domain experts involved, therefore, was the single remaining source of probabilistic information.

In Chapter 5, we discussed various methods from the field of decision analysis that were developed for the elicitation of probabilities from experts. As these methods have found widespread use in the construction of decision-analytic models, we decided to employ them in our efforts to elicit probabilities for the oesophagus network. We focused on the use of a probability scale for marking assessments, on different presentation formats for the probabilities to be assessed, and on the use of gambles. Before commenting on our experiences with these methods, we would like to emphasise that, prior to the construction of the oesophagus network, the domain experts had little or no acquaintance with expressing their knowledge and clinical experience in terms of probabilities.

The probability scale we used with our domain experts for the oesophagus network was a horizontal line with the three anchors 0, 50, and 100. We asked the domain experts to mark the

assessments for *all* conditional probabilities pertaining to a single node in the network, given a single conditioning context, on the same scale. For example, for the context of a polypoid, circular carcinoma of more than 10 centimeters, the experts were asked to mark their assessments for the probabilities of the passage of solid food, for the passage of pureed food at best, of liquid food, and of no passage at all; the experts thus had to indicate four assessments on a single scale. We chose to follow this procedure as it would allow the experts to compare and verify their assessments, thereby reducing the risk of overestimation. Contrary to expectation, however, the experts indicated that they felt quite uncomfortable working with the probability scale: it gave them ‘very little to go by’. The request to mark several assessments on a single line further appeared to introduce a bias towards aesthetically distributed marks; this spacing effect was also described in Chapter 5, page 118.

Another problem in our first elicitation effort turned out to be that the probabilities to be assessed for the oesophagus network were communicated to the domain experts in mathematical notation as described on page 115. Our experts experienced considerable difficulty understanding conditional probabilities in this presentation format. Especially the meaning of what is represented on either side of the conditional bar appeared to be confusing. As a result, the experts had difficulties constructing a mental model of the situation referred to and could not focus on just the assessment task at hand.

The frequency format, described on page 116, is generally easier to understand for experts than the mathematical notation and has been reported to be less liable to biases. With the frequency method, the domain experts were asked to envisage a population of one hundred patients suffering from an oesophageal carcinoma with certain characteristics. They were asked to assess the number of patients from among this population who would show a characteristic under study. Unfortunately, our experts had difficulties visualising the numbers of patients mentioned in the fragments of text: since oesophageal carcinoma has a low incidence in the Netherlands, visualising one hundred patients with a certain combination of characteristics turned out to be a demanding, if not impossible, task.

The use of lotteries for probability elicitation (see Section 5.3.2), was unfortunately also hampered by several difficulties. The experts indicated that they often felt that the lotteries were very hard to conceive because of the rare or unethical situations they represented. Moreover, the use of lotteries tended to take so much time that it soon became apparent that the elicitation of several thousands of conditional probabilities was quite infeasible in this way.

7.2 Evaluation of the elicitation method

In this section, we evaluate the use of our new method for probability elicitation. We will comment upon the observations made by the domain experts and the elicitors involved.

7.2.1 Using the method

In the first interview with our two domain experts, we informed them of the basic ideas underlying the new elicitation method. The general format of the fragments of text was demonstrated and the intended use of the response scale was detailed. We explained the way in which the

fragments of text and associated scales were grouped and instructed the experts to simultaneously take into consideration the probabilities from the same conditional probability distribution by spreading out in front of them the various sheets of paper pertaining to these probabilities. Finally, we explained to the experts that their probability assessments would be subjected to a sensitivity analysis that would reveal the sensitivity of the network's behaviour to the various assessments, and that, if necessary, we would try to refine the most influential ones later on. The basic idea of sensitivity analysis was explained in some detail to reassure the experts that rough assessments for the requested conditional probabilities would suffice at this stage in the construction of the network.

The elicitation of all conditional probabilities required for the diagnostic part of the oesophagus network outlined in Appendix A took five interviews of approximately two hours each over a period of fifteen months. Each interview focused on a small coherent part of the network. Prior to each interview, the elicitors spent some ten hours preparing the fragments of text and associated response scales to be presented to the experts. After the interview, it took the elicitors two to five hours to process the obtained assessments. The new method allowed the domain experts to give their assessments at a rate of 150 to 175 probabilities per hour; the remaining time was spent on explanation and instruction.

In the last interview, the domain experts were asked to evaluate the use of the new method of probability elicitation. For this purpose, we prepared a written evaluation form so as to not influence their observations. A translated version of this evaluation form is presented in Appendix C. The domain experts were asked whether or not the different ingredients in the method had helped them in the assessment task. Also, we asked for their opinion of the specific anchors on the response scale. The domain experts indicated that overall they had felt very comfortable with the method. They found the method most effective and much easier to use than any method for probability elicitation they had been subjected to before. Before commenting on their observations in more detail, we would like to point out that during the earlier, rather unsuccessful elicitation efforts, our domain experts had acquired some proficiency in expressing their knowledge and personal clinical experience in probabilities. As a result, they now appeared less daunted by the assessment task.

We recall from Chapter 6 that one of the ideas underlying our elicitation method is the use of a fragment of text, stated in terms of likelihood, to communicate a conditional probability to be assessed to the domain experts. During the interviews the elicitors had noticed that these fragments of text worked very well as additional explanation of the requested probabilities was seldom necessary. The two domain experts confirmed this observation and indicated that they had had no difficulties understanding the described probabilities. The elicitors had further noted that the characteristics described in the fragments of text served to call to mind specific patients or cases from scientific papers. Although the experts could not visualise a large group of patients with certain specific characteristics, their extensive clinical experience with cancer patients in general and their knowledge of reactive growth of cancer cells, along with information recalled from literature, enabled them to provide the required assessments without much difficulty.

With respect to the response scale used for marking assessments, the domain experts indicated that they had found the presence of both numerical and verbal anchors quite helpful. They mentioned that, when thinking about a conditional probability to be assessed, they had

used words as well as numbers. Depending on how familiar they felt with the characteristics described in the fragment of text, they preferred using the verbal or numerical expressions for marking their assessment on the scale. For example, the more uncertain they were about the probability to be assessed, the more they were inclined to think in terms of words. The verbal anchors on the scale then helped them to determine the position that they felt expressed the probability they had in mind. The elicitors noticed in the consecutive interviews that it became more and more easy for the experts to express their assessments as numbers. In the first few interviews they often stated a verbal expression and then encircled the appropriate anchor or put a mark close to the anchor on the scale. In the later interviews, they considered the entire response scale, marked their assessment, and subsequently wrote a number next to their mark.

The two domain experts further mentioned that they had felt comfortable with the specific verbal anchors used on the response scale. They indicated, however, that the expression “impossible” is hardly ever used in oncology. Especially in their communication with patients, oncologists seem to prefer the more cautious expression “improbable” to refer to almost impossible events. As a consequence, our domain experts tended to interpret the expression “improbable” as a 5% or even smaller probability rather than as a probability of around 15%. However, since the response scale provided both words and numbers, they had no difficulty indicating what they meant to express. The experts also mentioned that an extra anchor for 40% would have been useful. Note that these observations pertain to the lower half of the scale only. We would like to add that our response scale hardly accommodates for indicating extreme probability assessments, that is, assessments very close to 0% or 100%. There are no anchors close to zero and one hundred percent probability since only very few subjects in the first study from Chapter 6 generated extreme verbal expressions. The domain experts never seemed to want to express such extreme assessments either. When asked about this, they confirmed the correctness of our observation.

Another ingredient of our method is the grouping of the fragments of text in such a way that the probabilities from the same conditional distribution are taken into consideration simultaneously. As mentioned before, the domain experts were advised to spread out on the table in front of them the various sheets of paper pertaining to these probabilities. They were encouraged to focus first on the probabilities from a conditional distribution that were the easiest to assess, and then to use these as anchors for distributing the remaining probability mass over the more difficult ones. This turned out to be a most effective heuristic for eliciting assessments for nodes with more than two or three values. Especially in later interviews, the domain experts were able to verify the coherence of their assessments for the same conditional distribution without help, and adjusted them whenever they thought fit.

7.2.2 The use of trends

During the elicitation interviews with our domain experts, the concept of *trend* emerged. We use the term ‘trend’ to denote a fixed relation between two conditional probability distributions. To illustrate the concept of trend, we address the node *Invasion* that models the depth of invasion of an oesophageal carcinoma into the wall of a patient’s oesophagus. This node can take one of the values $T1$, $T2$, $T3$, and $T4$; the higher the number indicated in the value, the deeper the carcinoma has invaded into the oesophageal wall and the worse the patient’s prognosis. For the node *Inva-*

sion, several conditional probabilities were required, pertaining to differing shapes and varying lengths of the carcinoma. Upon assessing these probabilities, the domain experts started with the probabilities for the depth of invasion of a *polypoid* oesophageal carcinoma with a length of less than 5 centimeters. They subsequently indicated that patients with *ulcerating* tumours of this length were 10% worse off with regard to the depth of invasion of the carcinoma than patients with similar polypoid tumours. They thus explicitly related two conditional probability distributions to one another. As trends appeared to be a quite natural way of expressing probabilistic information, we encouraged our domain experts to provide trends wherever appropriate.

We designed a generic method for dealing, in an intuitively appealing and mathematically correct way, with the trends provided by our domain experts. The method is best explained in terms of the example trend given above. Suppose that, given a *polypoid* oesophageal carcinoma of less than 5 centimeters in length, the probabilities for the four different values of the node *Invasion* are assessed at x_1, x_2, x_3 , and x_4 — x_i being the probability assessment for the value T_i . The probabilities $x_i, i = 1, 2, 3, 4$, constitute the *anchor distribution* that is to be adjusted by the indicated trend to compute the probabilities for the related distribution. After consultation with our domain experts, we interpreted the specified trend as follows: *10% of the patients with a polypoid tumour of less than 5 centimeters with T_i for its depth of invasion would have had T_{i+1} for the depth of invasion if the tumour would have been an ulcerating tumour, $i = 1, 2, 3$.* The basic idea of the interpretation of the trend is depicted in Figure 7.1. For the probability

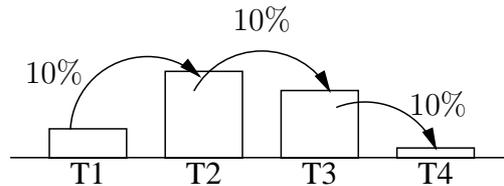


Figure 7.1: An schematic representation of handling trends.

assessments y_1, y_2, y_3 , and y_4 for the different values of the node *Invasion* given an *ulcerating* oesophageal carcinoma of less than 5 centimeters, we find

$$\begin{aligned} y_1 &\leftarrow x_1 - 0.10 \cdot x_1 \\ y_2 &\leftarrow x_2 - 0.10 \cdot x_2 + 0.10 \cdot x_1 \\ y_3 &\leftarrow x_3 - 0.10 \cdot x_3 + 0.10 \cdot x_2 \\ y_4 &\leftarrow x_4 + 0.10 \cdot x_3 \end{aligned}$$

It is readily verified that y_1, y_2, y_3 , and y_4 lie between 0 and 1, and together sum up to 1. In addition, it will be evident that this method for handling trends can easily be generalised to nodes with an arbitrary number of values and to trends specifying other percentages and other directions of adjustment.

7.3 Evaluation of the elicited probabilities

To assess the quality of the probabilities obtained with our new elicitation method, we conducted an evaluation study of the oesophagus network. In the study, we used data from patients from the

Antoni van Leeuwenhoekhuis diagnosed with oesophageal carcinoma. In Section 7.3.1, we analyse the probabilities obtained; we compare them with the data in Section 7.3.2. In Section 7.3.3, we study the probabilities in the context of the network. For this purpose, we entered for each patient, all diagnostic symptoms and test results available and computed the most likely stage of the patient's carcinoma from the network; we then compared the computed stage with the stage recorded in the data.

7.3.1 The obtained probabilities

When we set out to quantify the diagnostic part of the oesophagus network it included 39 nodes, requiring a total of 900 probabilities. The number of probabilities to be assessed per node ranged from 3 to 144, constituting a total of 267 (conditional) probability distributions. Many of the assessments we obtained from our domain experts equalled either 0 or 1: the experts gave 312 zeroes and 100 ones, together amounting to 46% of the network's probabilities. We would like to note, however, that 144 of these probabilities pertain to the deterministic node that models a carcinoma's stage, that is, 35% of the zeroes and ones constitute the (degenerate) conditional probability distributions for a *single* node. The domain experts further specified many probabilities on the lower half of the response scale: 72% of the assessments were less than or equal to 0.50.

For 12 of the 39 nodes from the network, the domain experts indicated trends, as discussed in the previous section. Using these trends, 241 probabilities were computed from other assessments. Of the total of 900 probabilities, therefore, 73% were assessed directly and 27% indirectly by adjustment of other probabilities. The indirect assessments pertained to 65 different conditional probability distributions. The trends indicated by the domain experts ranged from *equal* to the anchor distribution to a *20% shift*, in either direction, from this distribution.

To study the overall distribution of the assessments obtained with our elicitation method, we performed a frequency count. Figure 7.2(a) summarises the frequencies of all assessments obtained, be it directly or indirectly; we restricted the figure to the assessments not equal to zero or one. Figure 7.2(b) shows the frequencies of the assessments that were specified directly by the domain experts; once again we excluded zero and one from the figure. The two tables from Figure 7.3 show the ten most frequently specified assessments, counted over all assessments and over the direct assessments only.

We recall from Chapter 6 that the response scale used with our elicitation method specifies seven numerical anchors: 0, 15, 25, 50, 75, 85, and 100, or 0, 0.15, 0.25, 0.50, 0.75, 0.85, and 1.00, alternatively. By comparing our experts' assessments with these anchors, we find that 54% of all assessments and 63% of all direct assessments coincide with anchors. Focusing on the non-extreme assessments, that is, excluding 0 and 1, we find that 16% of all assessments and 20% of the direct assessments are anchors. The frequency counts from Figure 7.3 further reveal that among the ten most often specified assessments, there are four anchors from the response scale: 0, 0.15, 0.85, and 1.00. Among the ten most frequently specified direct assessments, there even are six anchors: 0, 0.15, 0.25, 0.75, 0.85, and 1.00. These findings are consistent with the often reported observation that the external stimulus used, in our case the response scale, plays a dominant role in the elicitation process. To conclude our discussion of the probabilities

obtained, we observe that, while the experts indicated that an extra anchor for 0.40 would have been helpful, they gave this assessment only seven times.

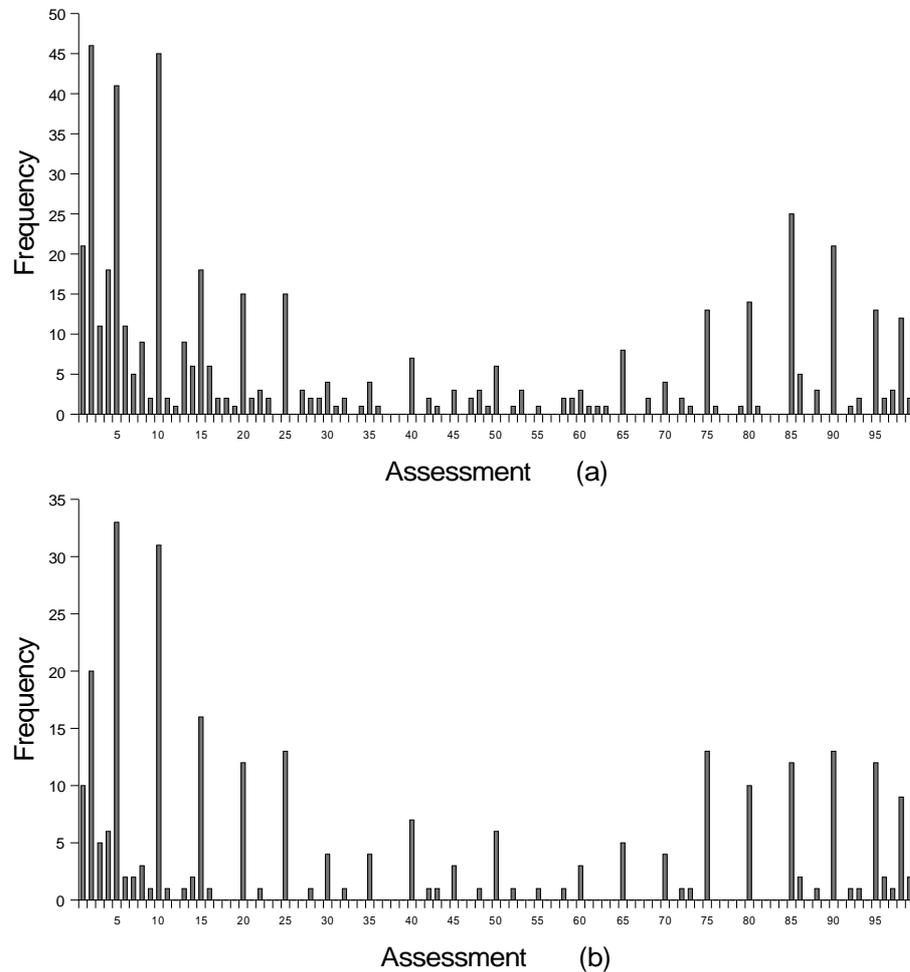


Figure 7.2: The distribution of all assessments obtained, (a), and of the assessments that were specified directly, (b), zeroes and ones excluded.

7.3.2 A comparison with the data

As described in Section 7.1, we had not been able to compose a large and rich enough data collection to allow for reliable assessment of the probabilities required for the oesophagus network. Our efforts to compose such a data collection, however, had resulted in data from historical records of 185 patients diagnosed with oesophageal cancer from the Antoni van Leeuwenhoekhuis. As these data had not been used for probability assessment, we could now exploit them for evaluation purposes. In this section, we compare the probabilities given by our domain experts with estimates from these data. Before doing so, however, we would like to note that the data collection does not constitute a fully independent source of information, as the collection consists of data from patients treated by our domain experts. Since the historical records dated

<i>assessment</i>	<i>frequency</i>	<i>assessment</i>	<i>frequency</i>
0	312	0	272
1.00	100	1.00	92
0.02	46	0.05	33
0.10	45	0.10	31
0.05	41	0.02	20
0.85	25	0.15	16
0.01	21	0.25	13
0.90	21	0.75	13
0.04	18	0.90	13
0.15	18	0.85	12

(a) (b)

Figure 7.3: The ten most frequent assessments, (a), and the ten most frequent direct assessments, (b).

back to between 1978 and 1985, and the experts did not scrutinise the data prior to assessing the required probabilities, we concluded that the data were independent enough to render the evaluation results meaningful.

We estimated, from our data collection, as many probabilities for the oesophagus network as possible. For only 26 of the 39 nodes involved, however, probabilities estimates could be computed: the remaining 13 nodes were not recorded in the data. Furthermore, for the nodes that were recorded, not all probabilities required could be estimated, as several combinations of values were missing in the data collection. The data provided for the estimation of 368, or 41%, of the network's probabilities, pertaining to 125 conditional distributions.

To investigate whether or not the probability assessments provided by our domain experts matched the estimates that we obtained from the data, we computed a *95% confidence interval* for each of the 368 probability estimates. The 95% confidence interval of a specific estimate is the interval in which the 'true' probability lies with 95% certainty; the length of the confidence interval thus quantifies the uncertainty in the estimate. The most common equation for computing a 95% confidence interval for a probability estimate p is the Wald approximation:

$$p \pm (1.96 \cdot \sqrt{\frac{p \cdot (1 - p)}{n}}),$$

where n is the number of patients whose data were used in the computation of the estimate p . Note that the larger the number of patients on which the estimate is based, the smaller the estimate's 95% confidence interval. The confidence intervals that we thus obtained for our probability estimates were rather large as a result of data sparseness: the intervals had an average length of 0.25. For 250 of the 368 estimates, the 95% confidence interval included the assessment that we had elicited from the experts. So, from the assessments that could be compared with the data, 68% more or less matched the probability estimates computed from the data.

As discussed before, our domain experts had indicated trends for 12 nodes, pertaining to 65 different conditional probability distributions. For 23 of these 65 trends, we could compare the

probabilities from both the specified anchor distribution and the distribution computed from the anchor, with probability estimates from the data. To determine the goodness of fit of a specific estimated distribution on the same distribution specified by the experts, we conducted a number of χ^2 -tests. The χ^2 -test is described on page 181 of Appendix B. To compare an estimated distribution with the same distribution specified by the experts, we took as *observed frequencies* the probabilities specified for the different values in the estimated distribution, and as *expected frequencies* the different probabilities from the distribution specified by the experts. Figure 7.4 summarises the results of the various χ^2 -tests that we conducted.

	<i>anchor distribution</i>	<i>computed distribution</i>	<i>both distributions</i>
<i>match</i>	15	13	8
<i>no match</i>	8	10	3

Figure 7.4: The number of matching anchor and indirectly computed distributions.

For 15, or 65%, of the 23 trends the anchor distribution given by the experts did not significantly differ ($\alpha = 5\%$) from the same distribution estimated from the data. Also, for eight of these 15 trends, the probability distribution that was computed from the anchor distribution by adjustment did not significantly differ from the distribution estimated from the data. For 35% of the trends specified by the experts, therefore, both the anchor distribution and the computed distribution closely matched the data. Of the eight trends of which the anchor distribution given by the experts differed significantly from the distribution estimated from the data, we found for three of them that also the computed distribution did not match the data. For 13% of the trends, therefore, both the anchor distribution and the computed distribution differed significantly from the distributions estimated from the data.

For the eight trends of which both the anchor distribution and the computed distribution closely matched the data, we may conclude that the direction as well as the percentage of adjustment that were indicated by our domain experts are correct. For the three trends of which both the anchor distribution and the computed distribution did not match the data, we investigated whether or not the specified trend was correct. For this purpose, we applied the trend, not to the anchor distribution given by the experts, but to the same distribution estimated from the data. For one of these trends, the thus computed probability distribution closely matched the data. We conclude that for a total of 9 trends, that is, for 39% of the trends specified by the domain experts, the indicated direction and percentage of adjustment are correct. Alternatively, 61% of the trends appear to be incorrect. Upon examining the fourteen apparently incorrect trends, we found that for four of them the basic idea underlying the trend seemed to be reflected in the data: for either an opposite direction or a weaker percentage of adjustment, the computed distribution matched the data. We would like to note that for many of the trends given by our experts only very few patient data were available as a basis for comparison. As a consequence, no conclusive statements with regard to the correctness of the specified trends can be made.

7.3.3 The quality of the network

To conclude our evaluation of the elicited probabilities, we conducted a study of the oesophagus network with data from 185 patients diagnosed with oesophageal carcinoma from the Antoni van Leeuwenhoekhuis. The study once again focused on the diagnostic part of the network that provides for establishing the stage of a patient's carcinoma; the stage of an oesophageal carcinoma can be either I, IIA, IIB, III, IVA, or IVB, in the order of progressive disease. Unfortunately, for 29 patients from our data collection the stage of their carcinoma was not recorded, leaving us with 156 patients for evaluation.

In a first evaluation of the oesophagus network, we entered for each patient from the data collection all diagnostic symptoms and test results available. We then computed the most likely stage of the patient's carcinoma from the network and compared it with the stage recorded in the data. Table 7.1 shows the results from this first evaluation. For 80 of the 156 patients, the stage of the carcinoma recorded in the data matched the stage that was computed from the network to have the highest probability. Assuming that the stages recorded in the data are correct, we concluded that the network established the correct stage for 51% of the patients. We would like to note that it is not uncommon to find a percentage in this range in initial evaluations of knowledge-based systems [7].

		<i>network</i>						
		I	IIA	IIB	III	IVA	IVB	<i>total</i>
<i>data</i>	I	2	0	0	0	0	0	2
	IIA	0	34	0	3	0	0	37
	IIB	0	3	0	3	0	0	6
	III	1	16	1	24	1	1	44
	IVA	1	9	2	23	6	1	42
	IVB	0	2	0	8	1	14	25
<i>total</i>		4	64	3	61	8	16	156

Table 7.1: The results from the first evaluation.

Taking the results from the first evaluation as a point of departure, we carefully examined the data of the patients for whom the probabilistic network returned a stage different from the recorded one. We identified three major sources of mismatch which could largely be attributed to the data. For 10 patients, the stage recorded in the data was acknowledged by the domain experts to be incorrect on retrospection. Various anomalies in the data constituted the second source of mismatch. For example, for some patients a deeper invasion of the carcinoma into the oesophageal wall was found during surgery than conjectured from endosonographic findings. For these patients, the *pre-surgical* findings and the *post-surgical* stage were recorded in the data. Because only the findings had been entered into the network, a stage different from the recorded one was established. The third major source of mismatch was found in the way findings had been entered into the patients' medical records. Often no distinction was made between facts and findings from diagnostic tests. For example, for many patients the medical record stated the presence of metastases in the cervical lymph nodes without indicating how this fact

had been established. Without explicitly stated test results the network could not establish the presence of these metastases, which resulted in an incorrect stage. The network so far included a single diagnostic test for establishing the presence or absence of metastases near the truncus coeliacus. This diagnostic test, a laparoscopic procedure, is rather invasive and has only recently been introduced into clinical practice. As it was very unlikely that this test had been performed for the majority of the patients from our data collection, we concluded that some nodes modelling diagnostic tests were missing from the network.

Building upon the above observations, we decided to perform a second evaluation of the network. For this purpose, we first extended the network with three extra nodes pertaining to diagnostic tests. In close consultation with our domain experts, we had identified two additional nodes for establishing the presence of metastases in the lymph nodes near the truncus coeliacus and one for establishing the presence of lymphatic metastases in the neck. In addition, we corrected the erroneous stages in the data, that is, as far as they had been identified in the first evaluation of the network.

In the second evaluation of the oesophagus network, we entered for each patient the available symptoms and test results, as before. If no tests were explicitly specified for facts with regard to lymphatic metastases in the neck or near the truncus coeliacus, we entered these facts as test results for the newly included nodes. In addition, we entered for each patient the facts stated in the data for which an indication of the performed test was missing; on average, 0.4 additional facts were entered per patient. The overall results of the second evaluation are shown in Table 7.2. Figure 7.5 summarises the results per stage. Figure 7.5(a) shows, per stage from the data, the percentage of patients for whom the network computed the same stage; these percentages can be interpreted as the sensitivity per stage of our network to the patient data. Figure 7.5(b) shows, per stage computed from the network, the percentage of patients for whom the data records the same stage; these percentages constitute the predictive value per stage of the network's outcome. Table 7.2 reveals that for 132 of the 156 patients, the stage of the carcinoma recorded in the (modified) data matched the stage computed from the network. Again assuming that the stages recorded in the data are correct, the network established the correct stage for 85% of the patients.

		<i>network</i>						
		I	IIA	IIB	III	IVA	IVB	<i>total</i>
<i>data</i>	I	2	0	0	0	0	0	2
	IIA	0	37	0	1	0	0	38
	IIB	0	1	0	3	0	0	4
	III	1	11	0	35	0	0	47
	IVA	0	0	0	4	35	0	39
	IVB	0	0	0	3	0	23	26
		3	49	0	46	35	23	156

Table 7.2: The results from the second evaluation study.

<i>stage from data</i>	<i>matched by network</i>	<i>stage from network</i>	<i>matched by data</i>
I	100%	I	67%
IIA	97%	IIA	76%
IIB	0%	IIB	–
III	74%	III	76%
IVA	90%	IVA	100%
IVB	88%	IVB	100%

(a)

(b)

Figure 7.5: The detailed results from the second evaluation.

7.4 Concluding observations

We used our new elicitation method for eliciting the probabilities required for the oesophagus network and evaluated its use with the domain experts involved. The experts indicated that they found the method much easier to use than any method for probability elicitation they had been subjected to before. Moreover, the method allowed the domain experts to give their assessments at a rate of over 150 probabilities per hour.

Using data from 185 patients, we evaluated the oesophagus network. A first evaluation revealed various sources of mismatch between the stage of a patient's carcinoma as recorded in the data and the one computed from the network. To a large extent, the mismatches could be attributed to anomalies in the data. We feel that this is not uncommon in evaluation studies like the present one. In addition, the first evaluation served to identify a small number of nodes missing from the network. After correcting the anomalies in the data and providing for the missing nodes, we found that a correct stage was established by the network for 85% of the patients. Given that the probabilities used are rough initial assessments and that the patient data require further cleaning up, the results from the study are quite encouraging. We are currently investigating the network's ability to predict the outcome of treatment and we hope to report the results in the near future.

For the construction of the oesophagus network, our newly designed elicitation method meant a major breakthrough. Prior to the use of our method, we had spent over a year experimenting, on and off, with other methods for probability elicitation, without success. Using our elicitation method, the probabilities for a major part of the oesophagus network were elicited in just two months' time. Our method seems to be well suited for eliciting the large number of probabilities that are typically required for a realistic probabilistic network. Although our method tends to ask considerable time from the elicitors for preparing the interviews with the experts, we feel that the ease with which probabilities can subsequently be elicited with the method makes this time certainly well spent.

Conclusions

In this thesis we have focused on the quantification task involved in the construction of probabilistic networks. More specifically, we have studied the use of qualitative approaches in the quantification process. The objectives of this thesis have been twofold: to refine the basic formalism of qualitative probabilistic networks and to design an elicitation method that permits the use of both verbal and numerical probability expressions. A summary of and conclusions from our studies pertaining to these two objectives are briefly presented in Sections 8.1 and 8.2, respectively; more in-depth discussions can be found in the different preceding chapters. Section 8.3 summarises open ends and outlines general directions for future research.

8.1 A qualitative approach to probabilistic reasoning

We have adopted the framework of qualitative probabilistic networks as a qualitative approach to probabilistic reasoning. Qualitative probabilistic networks serve three major purposes. First, as qualitative probabilistic networks allow for reasoning with a probabilistic network in a purely qualitative way, they can play an important role during the construction of a probabilistic network: they can be used to investigate the robustness of the structure before the network is quantified. Second, in a qualitative probabilistic network influences between nodes are represented by mathematically defined signs; using the underlying definitions, these signs can be used as constraints on the possible probability distributions during later quantification. And third, we envision the use of qualitative probabilistic networks as an important tool for explanation of probabilistic networks. For these three purposes, it is important to have a formalism that is as expressive as possible.

In the basic formalism of qualitative probabilistic networks, influences are modelled at a very coarse level of detail: an influence is either positive, negative, zero or ambiguous. More specific information, such as the strength of influences, influences that hold only in certain contexts, influ-

ences that are non-monotonic in the values of either the nodes involved in the influence or a third node, cannot be expressed in the formalism. As a result, upon reasoning, any non-monotonic influence and any trade-off modelled in a network will cause spreading of uninformative results throughout large parts of the network. In Chapter 4, we proposed refinements for the framework of qualitative probabilistic networks that attempt to solve these problems.

In Chapter 4 we have mainly considered binary nodes, only briefly touching upon the subject of non-binary nodes. In Section 8.1.1 we summarise the refinements that attempt to tackle the shortcomings arising from the coarse level of representation detail associated with qualitative probabilistic networks including only binary nodes. In Section 8.1.2 the problems associated with non-binary nodes are addressed. Two real-life probabilistic networks are abstracted to qualitative probabilistic networks in Section 8.1.3. Finally, in Section 8.1.4, we will discuss what we have achieved with our proposed refinements.

8.1.1 Refinements for binary nodes

The lack of information about strength, context and non-monotonicity in the original framework of qualitative probabilistic networks, may give rise to unnecessarily weak results upon reasoning with a network. However, with binary nodes, non-monotonicity of an influence can only be caused by the values of a third node and not by the values of the nodes directly involved in the influence. In Chapter 4, we showed that explicitly distinguishing between non-monotonic and unknown signs allows for resolving the non-monotonicity. One way of determining the sign of a resolved non-monotonicity is with the help of additive synergies. Non-monotonic influences are in essence influences that are positive in one context and negative in another; an alternative for resolving non-monotonicity is therefore to specify different signs for an influence in different contexts. Context-specificity of influences, in addition, can help to reveal hidden zero influences and to specify influences with different strengths in different contexts. We have added a notion of strength to the original framework of qualitative probabilistic networks by introducing an additional level of detail. This allows us to distinguish between strong and weak influences, where a strong influence is any influence that is stronger than all weak influences.

The proposed refinements serve for modelling influences at a more fine-grained level of detail. As a result, fewer ‘?’s will be present in the network and, during reasoning, fewer ‘?’s will be generated as a result of trade-offs. If the proposed refinements do not resolve all trade-offs in a qualitative probabilistic network, the remaining trade-offs can be isolated from the network by the pivotal pruning algorithm presented in Section 4.4. The algorithm subsequently identifies all information that is necessary to resolve the trade-off.

The original sign-propagation algorithm is designed to investigate the effect of propagating a single observation in light of previous observations. In practice, however, one is often interested in the effect of multiple simultaneous observations. Section 4.5 presented an elegant algorithm that can propagate multiple observations without producing unnecessary uninformative results.

8.1.2 Refinements for non-binary nodes

In qualitative probabilistic networks including non-binary nodes, influences can be present that are non-monotonic in the values of the nodes directly involved in the influence. We will illustrate

this with an example. Suppose we have a qualitative influence between two nodes A and B having three values each. Let

$$\begin{aligned} \Pr(b_1 | a_1) &= 0.5 & \Pr(b_1 | a_2) &= 0.4 & \Pr(b_1 | a_3) &= 0.3 \\ \Pr(b_2 | a_1) &= 0.1 & \Pr(b_2 | a_2) &= 0.4 & \Pr(b_2 | a_3) &= 0.5 \\ \Pr(b_3 | a_1) &= 0.4 & \Pr(b_3 | a_2) &= 0.2 & \Pr(b_3 | a_3) &= 0.2 \end{aligned}$$

Now suppose that the orders of the values of the nodes are as follows: $b_1 > b_2 > b_3$ and $a_1 > a_2 > a_3$. Then, $\Pr(B \geq b_1 | a_i) \geq \Pr(B \geq b_1 | a_j)$ for all $a_i > a_j$, but for all $a_i > a_j$ we have that $\Pr(B \geq b_2 | a_i) \leq \Pr(B \geq b_2 | a_j)$. The influence of node A on node B is therefore non-monotonic and the non-monotonicity is due to the order of the values of B . The influence would be positive if $b_1 > b_3 > b_2$, however. Now suppose that $b_1 > b_3 > b_2$ and $a_2 > a_1 > a_3$, then $\Pr(B \geq b_i | a_2) \leq \Pr(B \geq b_i | a_1)$ for all b_i , but also for all b_i we have that $\Pr(B \geq b_i | a_1) \geq \Pr(B \geq b_i | a_3)$. The non-monotonicity of the influence of A on B is now due to the order of the values of A . Non-monotonicity of an influence can also be caused by the order of the values of both nodes involved.

The non-monotonicity just described can be easily resolved by changing the order of values. Changing the order of values on which a natural order exists, for example the values of a tumour's length, will lead to counter-intuitive results. In addition, changing the order of a node's values will have effect on the signs of all influences in which the node is involved and can therefore easily lead to new non-monotonic influences.

In addition to the presence of an additional form of non-monotonicity, qualitative probabilistic networks including non-binary nodes have the same shortcomings as networks with binary nodes only. In the discussion that concludes each of the sections from Chapter 4 in which we propose a refinement of the basic formalism of qualitative probabilistic networks, we have described how the refinement applies to networks with non-binary nodes. Extension to non-binary nodes is mostly straightforward using the dummy-value approach proposed in Chapter 3. For qualitative probabilistic networks enhanced with a notion of strength, however, it is still unclear how to extend the definition to the non-binary case. The most obvious option for defining a strongly positive influence between a node A and a node B ,

$$\Pr(B \geq b_i | a_i) - \Pr(B \geq b_i | a_j) \geq \alpha, \quad \forall b_i, a_i > a_j,$$

is not possible as $\Pr(B \geq b_0 | a_i) = 1$ for $b_0 < b_i, i \neq 0$ and all values a_i of A . Finding a suitable definition, therefore, still calls for further investigation.

8.1.3 Qualitative abstractions of probabilistic networks

Qualitative probabilistic networks can be used for reasoning with an, as yet, unquantified probabilistic network. Qualitative probabilistic reasoning allows for investigating the effects of changing the network's structure and for testing the structure's robustness before quantification is commenced. Qualitative probabilistic networks, however, have not yet been used for this purpose with real-life probabilistic networks. To give an impression of a qualitative probabilistic network for a real-life application, we computed qualitative abstractions of the well-known ALARM-network [6] and of the diagnostic part of the oesophagus network. We will compare

the difference in expressiveness of the networks, using just the original formalism of qualitative probabilistic networks, and using the formalism refined with a notion of strength, context and explicit non-monotonicity.

The ALARM-network is a probabilistic network that simulates an anaesthesia monitor. The acronym ALARM stands for ‘‘A Logical Alarm Reduction Mechanism’’. The ALARM-network consists of 37, mostly non-binary, nodes and 46 arcs. The number of influences associated with arcs in the network therefore equals 46. The oesophagus network is described in Appendix A.

	# regular influences where sign δ is:					total
	+	-	0	\sim	?	
ALARM	17	9	0	5	15	46
oesophagus	32	12	0	0	15	59

Table 8.1: The number of positive, negative, zero and ambiguous regular influences for the ALARM-network and the diagnostic part of the oesophagus network, respectively.

Table 8.1 summarises for both the ALARM-network and the oesophagus network, the number of influences that are positive, negative, zero or ambiguous according to the regular definition of qualitative influence. The two networks do not specify any unknown signs. Ambiguous influences are therefore non-monotonic in either their own values, indicated by a ‘?’ , or in the values of a third node, indicated by ‘ \sim ’.

Using context-specific signs, hidden zeroes can be revealed, as well as positive and negative influences underlying the ambiguous influences that are non-monotonic. To give an indication of the number of influences for which a context-specific sign provides additional information, we identify for each regular qualitative influence the number of maximal contexts for which the influence’s sign can be specified. For a regular qualitative influence $S^\delta(A, B)$ associated with the arc $A \rightarrow B$, we take this number of contexts to be 1 (the empty context) if node B has no parents $X = \pi(B) \setminus \{A\}$ other than node A ; otherwise the number of contexts is taken to be the number of possible combinations of values for the set of nodes X . With each context underlying a regular qualitative influence is associated a context-specific influence. For each set of regular qualitative influences with the same sign, Table 8.2 presents the total number of contexts c_π covered by the regular influences, and the number of contexts for which the context-specific sign is positive, negative, zero or still ambiguous. From this table we have, for example, that the 17 regular influences with sign $\delta = +$ from the ALARM network together cover 59 different contexts, of which 38 are positive and 21 are actually zero. We would like to note that we computed the non-monotonic influences and context-specific signs by hand; it is therefore not guaranteed that we have found all those present in the network.

From Table 8.1 we have that in the ALARM-network, 35% of the regular qualitative influences are positive, 17% negative, and 48% ambiguous. Distinguishing between influences that are non-monotonic in a third node’s values and influences that are otherwise ambiguous, we find that 32% of the ambiguous influences are actually non-monotonic. When using context-specific influences, Table 8.2 displays that 32% of the influences are positive, 31% negative, 20% zero, and 17% remain ambiguous. It is possible that a number of the influences that remain ambiguous can be resolved by changing the order of the values of the nodes involved. This has not been tried

ALARM	# c_π with sign:				total
δ :	+	-	0	?	
+	38	-	21	-	59
-	-	40	11	-	51
0	-	-	-	-	0
\sim	10	15	4	-	29
?	24	9	8	28	79
<i>total</i>	72	64	44	28	218

oesophagus	# c_π with sign:				total
δ :	+	-	0	?	
+	74	-	8	-	82
-	-	36	8	-	44
0	-	-	-	-	0
\sim	-	-	-	-	0
?	6	3	2	38	49
<i>total</i>	80	39	18	38	175

Table 8.2: The number of contexts c_π with positive, negative, zero and ambiguous context-specific influences, covered by the regular influences with positive, negative, zero or ambiguous sign δ , for the ALARM-network and the diagnostic part of the oesophagus network, respectively.

since the values of the nodes in the ALARM-network all have a natural ordering upon them.

For the oesophagus network, Table 8.1 shows that 54% of the regular qualitative influences are positive, 21% negative, and 25% ambiguous. From Table 8.2 we have that, using context-specific signs, 46% of the influences are positive, 22% negative, 10% zero, and 22% remain ambiguous. To obtain these results the order of the values of the nodes ‘Shape’ and ‘Invasion-organs’ (see Figure A.3) were changed into ‘scirrheus < polypoid < ulcerating’ and ‘none < mediastinum < diaphragm < heart < trachea’, respectively.

We observe that in the ALARM-network and in the oesophagus network, the use of context-specific signs serves to reveal a considerable number of zero influences, and to decrease the number of ambiguous influences. Adding a notion of context thus largely enhances the expressiveness of both the ALARM-network and the oesophagus network.

	$\alpha \in$					
	[0, 0.12]	(0.12, 0.25]	(0.25, 0.75]	(0.75, 0.83]	(0.83, 0.85]	(0.85, 1]
# $\delta = ++$	6	5	4	3	2	0

Table 8.3: The number of strongly positive influences for different cut-off values α in a small fragment of the oesophagus network.

We will now briefly describe the effects of adding a notion of strength. To this end, we investigate a small fragment of the oesophagus network, the subgraph with the node ‘Haemametas’ as root, consisting of binary nodes only. The fragment contains seven nodes and six arcs; with each arc is associated a positive qualitative influence. Table 8.3 summarises the number of influences that are strongly positive for different values for the cut-off value α . We note that choosing a cut-off value of around 0.80 results in a nice balance between the number of strong and weak influences.

By abstracting the ALARM-network and the oesophagus network to qualitative networks, we have given an impression of the expressiveness of our refined formalism for qualitative probabilistic networks. With respect to the use of qualitative probabilistic networks for actual trade-off resolution, it will be more interesting to use the *prognostic* part of the oesophagus network when

that is completed. The prognostic part models numerous trade-offs between the desirable effects of different therapies and their complications.

8.1.4 Evaluation of our achievements

We have proposed to adopt the framework of qualitative probabilistic networks as a qualitative approach in the quantification process for constructing probabilistic networks. More specifically, we have proposed to use qualitative probabilistic networks for testing the robustness of the structure of a probabilistic network before quantification, and for supplying constraints on the probability distributions required for quantification. To make the formalism of qualitative probabilistic networks as expressive as possible, we described several refinements.

As the refinements we have proposed are all mathematically well-founded, probabilistic reasoning with a qualitative probabilistic network incorporating these refinements will always yield correct results. Theoretically, our refinements allow for revealing information about the relationships between variables that is abstracted away in the original formalism. Due to the additional information now present in a qualitative probabilistic network, reasoning with a network will more often lead to informative results, thus enabling us to more effectively test the robustness of the network's structure. In addition, the extra information provides stronger constraints on the probability distributions required for the probabilistic network under construction.

Since we have started our research into refinements for the framework of probabilistic networks, we have not been actively involved in the construction of the qualitative part of a real-life probabilistic network from scratch. We have therefore not been able to test the usefulness of the refinements in practice, except in the construction of small toy examples. However, qualitative abstractions of real-life probabilistic networks, such as the two discussed in the previous section, have convinced us that real-life networks most certainly contain non-monotonic influences, hidden zero influences and, obviously, influences of different strengths. We therefore believe that the proposed refinements are necessary extensions required for effective qualitative reasoning with real-life networks. In retrospect, we also believe that qualitative reasoning with the oesophagus network would have served to identify a number of missing nodes and a number of modelling errors in the network's structure. These modelling errors now remained undetected until the probability elicitation phase, and some even until the network was evaluated with real-life data.

From the described experiences we conclude that the ability to reason with a probabilistic network before it is quantified is very useful. Qualitative probabilistic networks are a tool that allow for doing this; when applied to real-life probabilistic networks, our proposed refinements are indispensable.

8.2 A qualitative approach to probability elicitation

When the process of constructing a probabilistic network has reached the actual quantification phase and probabilities have to be elicited from experts, these experts can be accommodated by using a probability elicitation method. Such a method should be used as part of an elicitation process, as described in Chapter 5. Chapter 5 also discusses the various elicitation methods currently available, their advantages and their drawbacks. A problem with these standard methods

is that they may work well for eliciting a few probabilities, but when hundreds or thousands of probabilities are required, they become too time-consuming. Some of these standard methods are, in addition, very complicated to understand.

A probability elicitation method that is suitable for eliciting the large number of probabilities required for a real-life probabilistic network should be easy to understand, easy to use, and should allow for eliciting a large number of probabilities in little time. We have developed such a method. The method combines various well-known ideas, such as describing the probability to be elicited in a fragment of text instead of using a mathematical notation, grouping the probabilities from the same conditional probability distribution, and presenting the experts with a separate probability scale for marking each assessment. The probability scale we use, however, is unique. We conducted a study into the use of verbal probability expressions. The result of this study is a list of seven distinguishable, commonly used verbal probability expressions, that have a stable rank order over subjects. The distances between these expressions as indicated by subjects, were used to project the verbal expressions onto a numerical probability scale resulting in a response scale with both verbal and numerical anchors. This scale is used in our probability elicitation method. A description of the studies, the results and the method was given in Chapter 6.

While we are still in the process of conducting a systematic study into the use of our new elicitation method, the method has already been used with two experts in oncology from the Netherlands Cancer Institute/Antoni van Leeuwenhoekhuis for quantifying the oesophagus network. Chapter 7 describes our experiences, and those of the experts, with probability elicitation using our new method, compared to standard methods. The experts felt very comfortable using the new method and were able to assess over 170 probabilities per hour. We also performed an evaluation of the behaviour of the oesophagus network quantified with probabilities obtained with the new elicitation method. The evaluation shows that, even though the assessments are only initial rough assessments, the network establishes the correct diagnosis for 85% of the patients in the patient database. The next step in quantifying the oesophagus network will be to perform a sensitivity analysis and refine only the most influential probabilities. This will be done either by using a third expert and/or by using more elaborate probability elicitation methods.

We have developed an elicitation method that has provided a major breakthrough in probability elicitation with our domain experts and our domain of application. The results of evaluating the initial assessments obtained with our method are also quite encouraging. We conclude that, although the method requires further evaluation, our experiences so far are very promising.

8.3 Directions for future research

In this section, we will present some directions for further research into reasoning with uncertainty and decision making under uncertainty, along the same lines as the research presented in this thesis.

8.3.1 Reasoning under uncertainty

We briefly indicate some general directions for future research with respect to qualitative approaches to elicitation and explanation, and for the framework of qualitative probabilistic net-

works; more detailed directions for further research concerning this latter subject were given in the different sections of Chapter 4.

The framework of qualitative probabilistic networks The refinements we proposed for the formalism of qualitative probabilistic networks still lack a definition for strong and weak influences in the non-binary case. In addition, in the first few sections of Chapter 4 we assumed that the sign-propagation algorithm is used for propagating only a single observation at a time. In that case, uninformative results arising from trade-offs are caused by trade-offs present in the network. In the last section of Chapter 4 we presented an extension of the sign-propagation algorithm that can be used for propagating multiple simultaneous observations. As a result of this extension, uninformative results can arise not only from trade-offs in the network, but also from conflicting observations. It will be interesting to investigate how the possibility to propagate multiple observations affects the other refinements proposed. For example, in qualitative probabilistic networks enhanced with a notion of strength, not only the strengths of influences are of importance, but also the strength or impact of the different observations will have to be taken into account.

We have taken the formalism of qualitative probabilistic networks introduced by Wellman and extended by Henrion and Druzdzel as a point of departure, taking the definitions of qualitative influence, additive synergy and product synergy for granted. It is however possible that still other definitions and other types of interaction can be found. In addition to refinements within the formalism, it would also be of interest to look into combining the formalism of qualitative probabilistic networks with other methods for qualitative reasoning. A number of more or less logical formalisms exist for reasoning with uncertainty in a qualitative way [47, 69, 93, 97]; these formalisms are however not designed for exploiting graphical structures of the domain under study. A first step towards combining such logical formalisms with a probabilistic network's graphical part has already been taken [88, 89].

Qualitative approaches and elicitation We have introduced qualitative probabilistic networks as a means of reasoning with a probabilistic network before it is actually quantified. Qualitative probabilistic networks, however, have not yet been used for this purpose with real-life networks. During the first attempts to quantify the oesophagus network, our experts were asked to indicate positive and negative signs for the arcs in a small part of the network. The signs seemed to come quite natural and the experts even produced 'double' signs on their own initiative. More experience with using a qualitative probabilistic network in the first step of quantifying a real-life probabilistic network is, however, required. When qualitative reasoning has led to the conclusion that the structure of a probabilistic network under construction can be considered robust, the actual probabilities are to be assessed. The qualitative relations from the qualitative network can be used as constraints on the possible probability distributions. Initial rough assessments for these probability distributions can be obtained with our, still to be systematically evaluated but promising, new elicitation method. A tool could be developed that includes the elicitation method as well as knowledge about a network's qualitative relationships. The constraints imposed by these relationships can be used as additional guidance for the experts during elicitation.

Qualitative approaches and explanation Qualitative probabilistic networks can be useful, not only during the construction of a probabilistic network, but also for explanation purposes. The ability of a probabilistic network to explain its output is perhaps the most important feature required for the acceptance of such networks. However, relatively little research has been done into the subject of explaining probabilistic networks. A discussion of the research that addressed this subject can be found in [118]. We believe that qualitative probabilistic networks, and in particular our pivotal pruning algorithm, can be used for explaining the reasoning process of a probabilistic network to its user. The actual output of a network can, in addition, be explained with the help of our response scale with both verbal and numerical anchors. The use of verbal probability expressions for explanation has been suggested before [32]. Research into the use of our response scale as an explanation tool is still required.

8.3.2 Decision making under uncertainty

Decision making under uncertainty is concerned with the problem of determining the best sequence of decisions to be made in the light of the uncertainties in the domain and the uncertain consequences of these decisions. Decision making under uncertainty involves both reasoning under uncertainty and reasoning about preferences. The framework of probabilistic networks, as discussed in this thesis, is tailored to reasoning under uncertainty and builds on the mathematical foundation of probability theory. The mathematical basis for reasoning about preferences is *utility theory*. To make the framework of probabilistic networks more suitable for decision support, it must incorporate not just uncertainty reasoning, but also reasoning about preferences and the possible decisions. As we therefore believe decision making to be a valuable addition to mere reasoning under uncertainty, we will briefly discuss a formalism for representing decision problems, the *influence diagram* [58].

Influence diagrams

An influence diagram can be seen as an extension of a probabilistic network, encoding not only a joint probability distribution, but also the various possible decisions that a decision maker can make and the desirability of the, uncertain, consequences of these decisions. As a probabilistic network, an influence diagram consists of a qualitative part and an associated quantitative part. The qualitative part of an influence diagram once again is an acyclic digraph. The set of nodes in the digraph is now partitioned into three different sets of nodes, having different meanings in the decision problem that is being represented. A node representing a domain variable is termed a *chance node*. A *decision node* models the various decision alternatives or actions that are at a decision maker's disposal; the value of a decision node is under direct control of the decision maker. The third type of node in the digraph is the *value node*, representing the desirability of the consequences that may arise from the various decisions alternatives; it may be looked upon as a real-valued, deterministic chance node. The value node is unique and does not have any outgoing arcs in the digraph.

The set of arcs in the digraph of an influence diagram is equally partitioned into different sets. The arcs between the chance nodes encode the independences among the represented variables. An arc from a decision node into a chance node expresses an influence on the chance node exerted

by the decision maker through his decision for the decision node at hand. The incoming arcs of a decision node together capture the information that is available at the time the decision is made. The digraph captures the basic assumption that, upon making a decision, all previously made decisions and previously available information are known to the decision maker. To conclude, an incoming arc of the value node expresses an influence on desirability.

The quantitative part of an influence diagram associates with each chance node a set of conditional probability distributions. In addition, with the value node is associated a *utility function*, that describes the desirability of each combination of values for the value node's parents. An influence diagram thus uniquely represents a decision problem. A solution to the represented problem is a decision or, in case of multiple decision nodes, a sequence of decisions that *maximizes* desirability of the consequences. To compute a solution, for each decision or sequence of decisions, the utilities of its various uncertain consequences are weighted with the probabilities that these consequences will occur. Efficient algorithms are available for computing from an influence diagram a solution for the represented decision problem, either by recursively reducing the diagram and combining probabilities and utilities [111], or by transforming an influence diagram into a probabilistic network and subsequently performing probabilistic inference [21].

In constructing an influence diagram the same problems are encountered as in the construction of a probabilistic network, and more: an influence diagram not only requires a large number of conditional probabilities, but also a possibly even larger set of utilities. While the prior and conditional probabilities in a network are fixed, the utilities will vary from user to user and thus have to be elicited for each user. For quantifying an influence diagram it is therefore again important that the structure of the diagram is considered robust, and that utility elicitation is made easy on the users. Another thesis can be filled with research into qualitative abstractions of influence diagrams and utility elicitation methods. Here, we will briefly discuss utility elicitation and *qualitative influence diagrams*.

Utility elicitation Utility elicitation has been studied extensively in the field of decision analysis [65, 129]. In fact, a number of probability elicitation methods described in Chapter 5 were originally designed for utility elicitation. There are two common approaches to utility elicitation [15]. The first is to base the utility functions on qualitative preferences elicited from the user. The second makes assumptions about the form and decomposability of the utility functions; decomposed functions are easier to understand and elicit.

Whereas utility elicitation methods were designed to elicit utilities for *decision trees*, they are now being applied for eliciting utilities for influence diagrams. Influence diagrams, however, tend to be much larger in size and, hence, require many more utilities. Utility elicitation using the standard methods thus becomes infeasible and other methods are called for. Recently, utility elicitation has started to receive attention in artificial intelligence. Attention has focused on the use of a combination of previously elicited utility functions and knowledge of the prevalence of these functions in the population of users for guiding the process of utility elicitation for a new user [15], and on the elicitation of partial utility models. For partial evaluation, an incremental approach was proposed [51], as well as an approach that includes uncertainty about a user's utilities [16]. The latter approach assumes a density function over possible utility values and applies statistical estimation techniques to learn such a density function from a database of

partially elicited utility functions. Focusing on eliciting only partial elicitation functions and applying techniques such as mentioned here may be the way to proceed. To investigate the effect of inaccuracies on the optimal decision, again sensitivity analysis can be used [23].

Qualitative influence diagrams *Qualitative influence diagrams* are qualitative abstractions of influence diagrams [134]. As a qualitative probabilistic network, a qualitative influence diagram bears a strong resemblance to its quantitative counterpart. It comprises the same graphical representation of the nodes involved in a decision problem, along with their different interrelationships. As with qualitative probabilistic networks, a qualitative influence diagram associates with the arcs between its chance nodes qualitative influences, additive synergies, and product synergies. As seen in Chapter 3, these qualitative relationships adhere to the properties of symmetry, transitivity, and parallel composition. Qualitative influences and synergies can also be specified between decision nodes and chance nodes. In addition, a qualitative influence diagram specifies various *qualitative preferential relationships* for the value node of the digraph. These preferential relationships are qualitative influences on utility and additive synergies on utility. The qualitative relationships together constitute a set of hyperarcs for the network's digraph.

For decision making in qualitative influence diagrams, an algorithm was designed based on the idea of recursively reducing a diagram [134]. However, this algorithm tends to create more ambiguities than necessary as a result of information loss during the reductions employed and, hence, is not able to compute a preferred decision in all situations where such a decision can be derived. In [101], we proposed a new, elegant algorithm, based on the sign propagation algorithm, that creates fewer unnecessary ambiguities and is therefore able to solve more decision problems. As a qualitative influence diagram embeds a qualitative probabilistic network for representing the qualitative relationships among its chance nodes, the original sign-propagation algorithm can be applied straightforwardly to the diagram's probabilistic part. In addition, the algorithm can be used to trace the effect of evidence on the value node, indicating the sign of change in expected utility. The sign-propagation algorithm, however, cannot be used for the decision nodes in the diagram, as it would ignore the control of the decision maker. To provide for these nodes, we adapted the algorithm so that it does not pass signs into decision nodes. The preferred decision alternatives are those that have a positive effect on expected utility. This is, if we should make a decision, enter this decision as evidence and use the sign propagation algorithm, then the value node should receive a '+'. To determine the decisions that have this effect, we work in the opposite direction and propagate a '+' from the value node towards every decision node, regardless of the sign of the influence on utility of the evidence. For more details on decision-making in qualitative influence diagrams, the reader is referred to [101].

The Oesophagus Network

The Netherlands Cancer Institute, Antoni van Leeuwenhoekhuis, in the Netherlands is a specialised center for the treatment of cancer patients. Every year some eighty patients receive treatment for *oesophageal carcinoma* at the center. These patients are currently assigned to a therapy by means of a standard protocol that includes a small number of prognostic factors. Based upon this protocol, 75% of the patients show a favourable response to the therapy instilled; one out of every four patients, however, develops serious complications as a result of the therapy. To arrive at a more fine-grained protocol with a higher favourable response rate, a decision-support system is being developed for patient-specific therapy selection for oesophageal carcinoma. The kernel of the system is a probabilistic network for diagnosis and prognostication; we will refer to this probabilistic network as the *oesophagus network*. The system is constructed and refined with the help of two experts in gastrointestinal oncology from the Netherlands Cancer Institute and is destined for use in clinical practice.

The illustrative examples used throughout this thesis, mostly comprise of simplified fragments from the oesophagus network. In addition, the network was used in a case study of our newly designed probability elicitation method described in Chapter 6; the results of the study, including an evaluation of the network, are provided in Chapter 7. Here, we will provide some background on the oesophagus network's structure, its probabilities, and the domain knowledge captured. The overall structure of our probabilistic network for oesophageal carcinoma is shown in Figure A.1.

Domain knowledge

A carcinoma may develop in a patient's oesophagus as a consequence of a lesion of the oesophageal wall, for example as a result of frequent reflux or associated with smoking and drinking habits. Due to the presence of an oesophageal carcinoma, a patient will often have difficulty

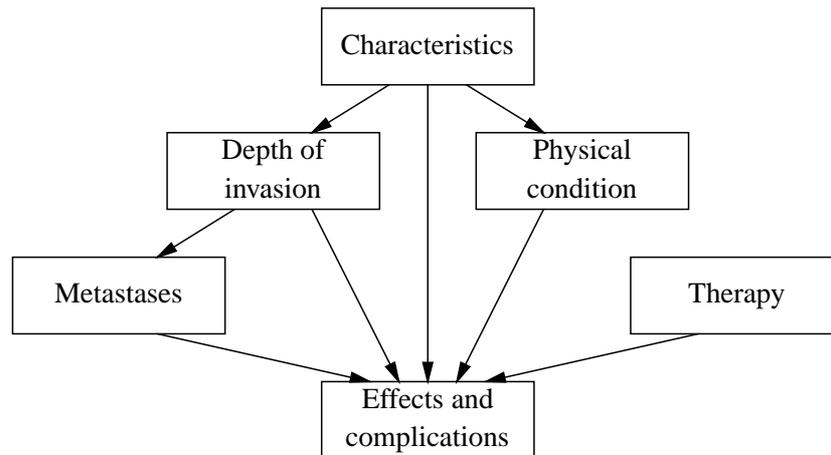


Figure A.1: The overall structure of the probabilistic network for oesophageal carcinoma.

swallowing food and may, as a consequence, lose weight. The extent to which a patient suffers from these complaints depends on the characteristics of the carcinoma, such as its location in the oesophagus and its histological type, length and macroscopic shape. These characteristics influence the carcinoma's prospective growth. An oesophageal carcinoma typically invades the oesophageal wall upon growth. When the carcinoma has grown through all three layers of the oesophageal wall, it may invade neighbouring structures such as the trachea and bronchi, the heart, the mediastinum, or the diaphragm, depending upon the location of the tumour in the oesophagus. In due time, the carcinoma may give rise to lymphatic metastases in, for example, the patient's cervical lymph nodes and to haematogenous metastases in, for example, the lungs and the liver. The characteristics, depth of invasion, and extent of metastasis are summarised in the carcinoma's stage; these factors, together with a patient's physical condition, largely influence a patient's life expectancy and are indicative of the effects and complications to be expected from the different therapeutic alternatives. To establish these factors in a patient, typically a number of diagnostic tests are performed, ranging from multiple biopsies of the primary tumour to gastroscopic and endosonographic examination of the oesophagus and a CT-scan of the patient's chest and liver. The tests differ considerably in their sensitivity and specificity characteristics. For example, for establishing the presence or absence of metastases in the loco-regional lymph nodes endosonography has a low sensitivity and specificity whereas gastroscopy for establishing the carcinoma's shape has considerably better sensitivity and specificity characteristics.

Whereas establishing the presence of an oesophageal carcinoma in a patient is relatively straightforward, the staging of the carcinoma and especially the selection of an appropriate therapy are far harder tasks. In the Antoni van Leeuwenhoekhuis, different therapeutic alternatives are available, ranging from surgical removal of the oesophagus, to radiotherapy and positioning a prosthesis in the oesophagus. The effects aimed at by instilling a therapy include removal or reduction of the patient's primary tumour to prolong life expectancy and an improved passage of food through the oesophagus. The therapies available differ in the extent to which these effects can be attained. For example, where the aim of surgical removal of the oesophagus is to achieve a better life expectancy for a patient, positioning a prosthesis in the oesophagus is aimed merely at relieving the patient's problems with swallowing food. Instillation of a therapy is expected

to be accompanied not only by beneficial effects but also by complications; these complications can be very serious and may in fact result in death. It will be evident that the possible effects and complications require careful balancing before a therapy is decided upon.

Description of the oesophagus network

The kernel of our decision-support system is a probabilistic network of oesophageal carcinoma, comprising of a diagnostic part and a prognostic part. The diagnostic part describes the various characteristics of an oesophageal carcinoma and the pathophysiological processes underlying its invasion into the oesophageal wall and its metastasis. This part of the model further captures the sensitivity and specificity characteristics of the diagnostic tests that are typically performed to assess a carcinoma's stage. The prognostic part describes seven possible therapies along with their possible effects and complications. This part further specifies the extent to which the effects and complications associated with each therapy influence life expectancy and the patient's ability to swallow food. When a patient's complaints and test results are entered, the diagnostic part of the network provides for establishing the stage of the patient's carcinoma; the prognostic part provides for subsequently predicting the most likely outcomes of the different treatment alternatives.

The probabilistic network of oesophageal carcinoma is being constructed and refined with the help of two domain experts. In a sequence of eleven interviews of two to four hours each, over a period of two years, the experts identified the relevant diagnostic and prognostic factors to be captured as nodes in the network, along with their possible values. The influential relationships between the nodes were elicited from the experts using the notion of causality: typical questions asked by the elicitors during the interviews were "What could cause this effect?" and "What manifestations could this cause have?". The thus elicited causal relationships were expressed in graphical terms by taking the direction of causality for directing the links between related nodes. Once the graphical structure of the network was considered robust, attention turned to the elicitation of the required probabilities. The author of this thesis was not involved in the network's construction until the second attempt at quantifying the network, using the newly designed elicitation method.

Our probabilistic network currently includes 66 nodes. Of these, 42 nodes pertain to the diagnostic part of the network and the remaining 24 nodes are used for prognostication. For the nodes, a total of 4009 (conditional) probabilities have been specified. For the purpose of this thesis, we will focus our attention on the diagnostic part of the network, which constitutes a coherent and self-supporting probabilistic network. The overall structure of the diagnostic part of the network is shown in Figure A.2; it is depicted in more detail in Figure A.3, showing the prior probability distribution per node. For each group of nodes from the diagnostic part, Table A.1 summarises the number of nodes, n , and the number of (conditional) probabilities, p , specified. The table, in addition, specifies for each group the mean, maximum, and minimum number of values v per node, number of incoming arcs i per node, and number of outgoing arcs o per node. The 42 nodes involved require 279 (conditional) probability distributions, with a total of 932 probabilities. The node requiring the largest number of conditional probability distributions, 24, and probabilities, 144, models the stage of the carcinoma; this node is a deterministic node. The non-deterministic node requiring the largest number of probability distributions, 20, is the

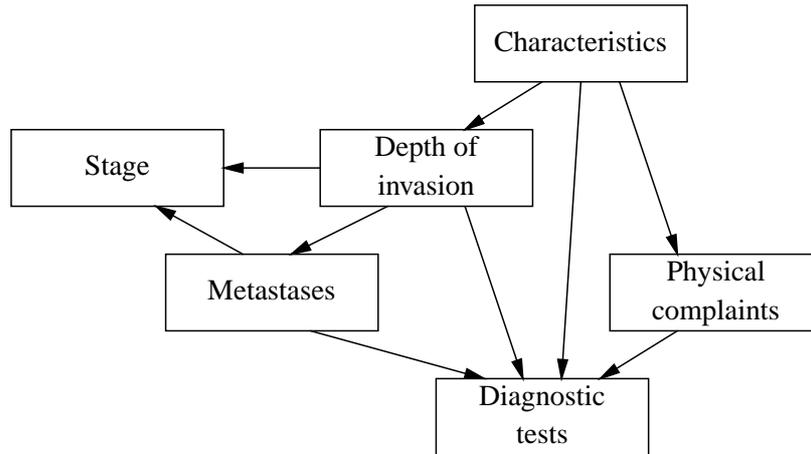


Figure A.2: The overall structure of the diagnostic part of the oesophagus network.

node describing the result of an endosonographic examination of a patient's mediastinum. The non-deterministic node requiring the largest number of probabilities is the node describing the result of an endosonogram of a patient's oesophagus with respect to the depth of invasion of the carcinoma in the oesophageal wall; it requires 80 probabilities.

	n	p	v			i			o		
			\bar{v}	max	min	\bar{i}	max	min	\bar{o}	max	min
Physical complaints	2	84	3.5	4	3	2	3	1	5.5	11	0
Diagnostic tests	23	470	2.7	5	2	1.4	2	1	0	0	0
Characteristics	5	24	2.8	3	2	0.4	1	0	3.2	5	1
Depth of invasion	4	140	3.3	5	2	2	2	2	3.5	6	1
Metastases	7	70	2.1	3	2	1.4	2	1	2.6	4	2
Stage	1	144	6	6	6	3	3	3	0	0	0
<i>total</i>	42	932	3.4	6	2	1.7	3	0	2.5	11	0

Table A.1: Some statistics concerning the number of nodes n , the number of (conditional) probabilities, p , the mean, maximum, and minimum number of values v per node, in-degree i per node and out-degree o per node, for the diagnostic part of the oesophagus network.

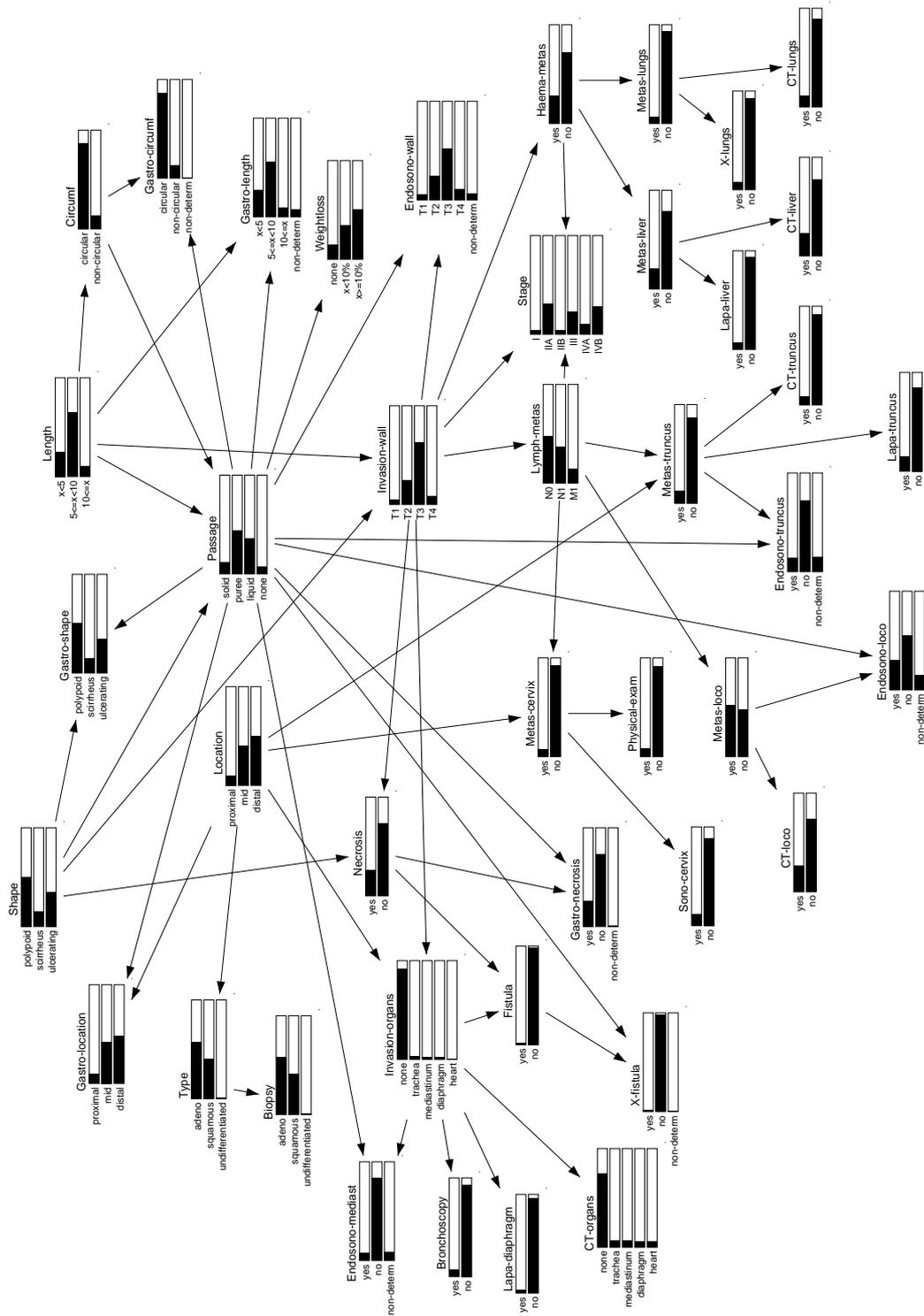


Figure A.3: The diagnostic part of the oesophagus network.

Statistical Methods

Analysis of variance (ANOVA)

ANOVA is an acronym for ANalysis Of VAriance. It is a method for testing for significant differences between mean values of a certain variable measured in samples from different populations, by analysing variance (SD^2). We briefly review the details necessary for understanding the results presented for the second study in Chapter 6. More details can be found in any textbook on statistical analyses (see e.g. [59]).

In the study concerned, we performed eight ANOVA analyses, one for each verbal probability expression. For each analysis, we considered samples from four populations: 26 ranking numbers provided by medical subjects in the no-context group, 26 ranking numbers provided by other subjects in the no-context group, 21 ranking numbers from medical subjects in the context group, and 22 ranking numbers from other subjects in the context group. We were interested in whether the mean ranking numbers of the samples differed between the populations.

ANOVA tests the null hypothesis that all population means are equal by comparing two estimates of variance. The Mean Square Error (MSE) estimate is an estimate of variance that captures variability of the data within the population samples. The Mean Square Between (MSB) estimate captures the variability between population samples. If the null hypothesis is true, then MSE and MSB should be approximately the same. To test the null hypothesis, therefore, the ratio $F = \frac{MSB}{MSE}$ is used. This ratio F has a probability distribution known as the *F-distribution*. From the *F-distribution*, the probability p of obtaining a value for F as large as or larger than the one calculated from the data can be determined. If p is lower than a pre-specified significance level α , then the ratio of the two variance estimates is significantly greater than 1 and the null hypothesis is rejected.

The function describing an F -distribution is defined by two parameters: the degrees of freedom of the numerator (df_n) and the degrees of freedom of the denominator (df_d), where the number of degrees of freedom is the number of parameters which may be independently varied. In the simplest case of ANOVA, we take

$$\begin{aligned}df_n &= \text{number of populations} - 1 \\df_d &= \text{overall sample size} - \text{number of populations}\end{aligned}$$

As our study included four populations with a total of 90 subjects providing one ranking number per expression, we are concerned with the $F(3, 86)$ -distribution. Using this distribution, ANOVA computed from our ranking number data for the expression **possible** the value 4.018 for F , which differs significantly from 1 ($p = .01$). We summarise the results for other expressions with mean ranking numbers differing significantly over groups: for **impossible** $F(3, 86) = 4.605$, $p = .005$, for **improbable** $F = 5.684$, $p = .001$, for **expected** $F = 4.481$, $p = .006$ and for **certain** $F = 4.390$, $p = .006$.

When the two estimates of variance differ significantly and the null hypothesis is rejected, then the only conclusion that can be drawn is that at least one population mean substantially deviates from at least one other population mean. ANOVA does not reveal which population means differ from which others. Upon finding significant variance differences, therefore, further analyses are usually conducted to investigate where the differences lie precisely. Such analyses are called post hoc tests. An example post hoc test is Tukey's HSD procedure for testing all pairwise comparisons of the means among the groups.

Principal component analysis (PRINCALS)

PRINCALS, an acronym for PRINCipal Components analysis by Alternating Least Squares, is a non-linear principal component analysis method developed at Leiden University. PRINCALS is capable of handling ordinal data. A principal component analysis describes a number of variables with a smaller number of variables, termed the principal components, that still contain as much information, exhibited in the original variables, as possible. We will briefly review PRINCALS; more details can be found in [45].

As an example, suppose we have ten different variables pertaining to restaurant characteristics. Suppose that half of these variables are concerned with the quality of the food and the service, and the other half concern the atmosphere in and around the restaurant. The information incorporated in these ten variables may therefore be summarised with only two variables. These two variables are then called the principal components or the *dimensions* of the problem. Suppose we would like to compare the quality of food and service, and the atmosphere for a number of restaurants, then analysing for all restaurants the summarised values of the principal components is much easier than analysing the values for all original variables. For each dimension and each observation object (the restaurants), PRINCALS computes a summarised value, called an *object score*; the total set of object scores is the *solution*. In addition, PRINCALS provides several measures of fit to indicate the quality of the solution, that is, how good the object scores represent the original variables' values. We will discuss two of these fitness measures. The *goodness-of-fit* of the entire solution is given by an *eigenvalue* (between 0.0000 and 1.0000)

for each dimension. The eigenvalue gives the proportion of variance in the original variables accounted for by the principal component. With two dimensions, a high eigenvalue on one dimension and a low eigenvalue on the other dimension indicates a one-dimensional solution. If the eigenvalues on both dimensions are high, this indicates at least a two-dimensional solution. PRINCALS also provides a measure of fit for each original variable, indicating for each dimension how good the object scores represent that variable's values; this fitness measure is termed the *component loading*.

In our second study in Chapter 6, subjects were asked to provide ranking numbers for eight verbal probability expressions. We were interested in comparing the ranking numbers for the expressions, summarised over all subjects. Therefore, the subjects in our study are the variables for the PRINCALS analysis and the eight expressions are the observation objects; each variable has a single value for each observation object, being the ranking number a subject has provided for that expression. If the subjects are all ordering the expressions from the same perspective, then the solution should be one-dimensional, that is, the rank orders of all subjects can be summarised in a single order. A high-quality solution will be found if all subjects have given comparable rank orders. If some of the subjects have ordered the expressions from a different perspective, then the data will contain more than one principle component; a two-dimensional solution, for example, would indicate that within two groups of subjects the orderings are comparable, but between the two groups they are not. From the component loadings provided for the subjects, it can be derived which subject is summarised in which principal component. The eigenvalues we found for the different PRINCALS analyses we described for our second study, are presented in Table B.1.

		<i>no-context</i>		<i>context</i>	
		group 1	group 2	group 3	group 4
<i>Analysis 1</i>	# subjects	26	26	21	22
	eigenvalue dim. 1	0.9675	0.9300	0.8182	0.6572
	eigenvalue dim. 2	–	–	0.1484	0.2986
<i>Analysis 2</i>	# subjects	24	25	12	11
	eigenvalue	0.9175	0.9500	0.9550	0.9515
<i>Analysis 3</i>	# subjects	49		23	
	eigenvalue	0.9504		0.9652	
<i>Analysis 4</i>	# subjects	72			
	eigenvalue	0.9658			

Table B.1: Eigenvalues found for the different analyses for the second study in Chapter 6.

Alternating least-squares scaling (ALSCAL)

ALSCAL is an acronym for Alternating Least-squares SCALing. It takes a set of distances between objects and “recreates the map” by computing the coordinates of the objects. Compare this with creating a map of Europe given the distances between each pair of cities. ALSCAL can compute coordinates in as many dimensions as necessary, in such a way that the Euclidean distances

between the coordinates will correspond as closely as possible to the originally provided object distances. As close as possible means that the difference (error) between the squared coordinate distances and the object distances is as small as possible. This error is provided by ALSCAL for its solution as a measure of fit; it is termed *s-stress* and ranges from 1 (worst possible fit) to 0 (perfect fit). Another fitness measure is *RSQ*, the squared correlation coefficient between the object distances and the coordinate distances; it ranges from 0 (worst) to 1 (best) and indicates the proportion of variance of the data that is accounted for by the coordinate distances. More detailed information on ALSCAL can be found in [119].

In our third study in Chapter 6, we constructed for each subject a matrix with distances between each pair of verbal probability expressions. We used a replicated version of ALSCAL to be able to compute a single solution for a set of matrices together. We computed one solution for the 26 medical subjects, one solution for the 52 other subjects, and one solution for the total group of 78 subjects. We treated our data as continuous ordinal data and as matrix-conditional, that is, we took into account that the meaning of the numbers in a matrix is conditional on the subject. Since our previous studies had indicated that all probability expressions seemed to be comparable, we did an analysis in only one dimension. The coordinates computed by ALSCAL for the eight expressions in the three groups can be found in Table 6.6; these coordinates are standardised to have a zero mean and a variance of one. Mean *s-stress* and *RSQ* values over all matrices in a group are presented in Table B.2.

	<i>group 1</i> (<i>n</i> = 26)	<i>group 2</i> (<i>n</i> = 52)	<i>all subjects</i> (<i>n</i> = 78)
<i>s-stress</i>	0.45425	0.40965	0.42827
<i>RSQ</i>	0.536	0.635	0.602

Table B.2: Fitness measures for the ALSCAL solutions for the subject groups from the third study in Chapter 6.

Log-Linear Analysis (HILOGLINEAR)

HILOGLINEAR is a log-linear analysis method. It is designed for analysing tables for three or more variables; the cells of the tables contain frequencies. A log-linear analysis is used to find a model of the dependencies between or interactions among variables. Here we present the details necessary for understanding the results described for the fourth study in Chapter 6. More details can be found in e.g. [28].

In our fourth study, we constructed $2 \times 2 \times 7$ matrices with the number of subjects that made a certain *Decision* for a situation described, using one of two presentation *Modes* for a probability *Expression*. These matrices can be represented by three-way tables with two rows and 14 columns. If no dependencies exist among the table's variables, then the frequencies in each cell of the table should proportionately reflect the total frequencies. For example, the proportion of 'yes' to 'no' decisions should be the same regardless of representation mode or expression, and should therefore equal the proportion of all subjects having answered 'yes' to all

subjects having answered ‘no’. Therefore, if all variables are independent, then

$$\text{expected cell frequency} = \frac{\text{column total} \times \text{row total}}{N},$$

where N is the total sum of frequencies. This can be written as

$$\log(\text{expected cell frequency}) = \log(\text{column total}) + \log(\text{row total}) - \log(N),$$

showing that by taking the logarithm, the expected cell frequency is a linear function of the other values involved. Such a model of expected cell frequencies reflecting certain (in)dependence assumptions is called a *log-linear model*.

The assumption of independence puts constraints on the expected cell frequencies. In fact, the above model is the most constrained model possible and, using the variables from our study, is denoted by **Expression + Decision + Mode**. Without any independence assumptions, there is no difference between expected frequencies and the actually observed frequencies and the model will always perfectly match the data. As a model matching any data gives no information about independences present in the data, this model is called informationless; it is denoted by **Expression \times Decision \times Mode**. Informally speaking, in a log-linear model, an ‘ \times ’ denotes (possible) dependence and a ‘+’ denotes independence. Any combination of variables with +s and \times s is a possible model. The model **Expression \times Decision + Mode**, for example, indicates that only the variables Expression and Decision are dependent and that Mode is independent of either.

Finding the log-linear model that best describes the data is a two-step process. The first step is testing the expected cell frequencies for all possible models for significant deviations from the data. All models for which expected cell frequencies and observed frequencies do not significantly differ are said to fit the data and are subsequently compared with each other to find the best one. To determine whether a model fits the data and what the goodness of this fit is, a χ^2 -test is performed. The null-hypothesis for this test is that the expected cell frequencies equal the observed frequencies. To test this hypothesis, the following measure is used:

$$\chi^2 = \sum_{\text{table cells}} \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

This measure has a χ^2 -distribution; the function describing the distribution is defined by a single parameter, the degrees of freedom (df). Here, we take df to be $(\#rows - 1) \times (\#columns - 1)$. From the χ^2 -distribution, the probability p of obtaining a value for χ^2 greater or equal to the one calculated from the data can be determined. If p is smaller than the significance level of 0.05, then the fit is too bad and the model should be rejected. The best model is the simplest, non-informationless, model with the highest p -value and the lowest value for χ^2 ; the smaller the χ^2 value, with a minimum of 0, the smaller the difference between expected and observed frequencies and therefore, the better the fit.

It is known that the χ^2 -tests are somewhat too liberal, finding differences more significant than they actually are; this is especially the case if some expected cell frequencies are small. It is therefore suggested to correct the observed frequencies before computing the fit [28]. This

continuity correction consists of adjusting each value of the observed frequencies by 0.25 towards its own expected frequency. In our study, we used this correction in situations where no simple model would fit the data and we thought this could be caused by small cell frequencies.

Table B.3 presents the χ^2 values computed for the results presented for our fourth study in Chapter 6. Separate results are given for the analysis of the number of subjects making a decision and the analysis of the total confidence they had in the correctness of the decision. In all cases, we eventually found **Expression \times Decision + Mode** to be the best fitting model; in most cases this model was found for the original data, in two cases it was found after a continuity correction and in one case it was found after deleting the probability expression represented by either 25% or **uncertain**. The first listed χ^2 in the two columns, is the value found for the data from the decision situation in which the probability expressions were rank ordered. The number of degrees of freedom (df) is 11 for the table with deleted data; $df = 13$ for all other tables.

	<i>decision tables</i>		<i>confidence tables</i>	
<i>original dataset</i>	$\chi^2 = 2.876$	$p = 0.998$	$\chi^2 = 2.574$	$p = 0.999$
	$\chi^2 = 15.468$	$p = 0.279$	$\chi^2 = 12.541$	$p = 0.484$
	$\chi^2 = 16.085$	$p = 0.245$	$\chi^2 = 7.520$	$p = 0.873$
	$\chi^2 = 9.438$	$p = 0.739$	$\chi^2 = 9.208$	$p = 0.757$
	$\chi^2 = 9.261$	$p = 0.753$		
<i>corrected dataset</i>	$\chi^2 = 20.342$	$p = 0.087$	$\chi^2 = 13.138$	$p = 0.437$
<i>reduced dataset</i>			$\chi^2 = 15.287$	$p = 0.170$

Table B.3: Goodness-of-fit for the results presented for the fourth study in Chapter 6.

Questionnaires and Evaluation Forms

On all questionnaires subjects were asked to provide their sex (m/f), year of birth and their profession or studies; they were then thanked for their cooperation. The following questionnaire was used for the first study of Chapter 6.

Expressing Uncertainty

People are used to coping with uncertain information when reasoning and making decisions. This information can contain uncertainty about the occurrences of certain events, about the consequences of events, or about the consequences of a decision, etc. Events are not simply “certain” or “uncertain”: we are capable of distinguishing degrees of uncertainty. In doing this, we often use verbal expressions as “possible”, “probable”, “uncertain”, etcetera. One can, for example, say “It is possible that I will pass my exam”, or “It is highly likely that I will go to Amsterdam tomorrow”. With this experiment we want to find out what verbal expression are commonly used.

We would like you to write down a list of expressions that, in your opinion, could be applied in situations in which you want to express a degree of uncertainty. To help you with this, we present a few example sentences:

- It is that it will rain tomorrow.
- It is that there will be a railway strike next week.
- It is that I will find a job this summer.

Please write down all expressions that you think are *commonly used* for expressing a degree of uncertainty and that are meaningful in the above contexts. Do not restrict yourself to terms that express what you expect will actually happen.

The following questionnaire was used in the second study of Chapter 6. We only present the context version; for the no-context version the probability expressions are presented in isolation in the order: possible, impossible, uncertain, certain, probable, improbable, expected, and undecided.

Ordering Verbal Probability Expressions

People are used to coping with uncertain information when reasoning and making decisions. This information can contain uncertainty about the occurrences of events, about the consequences of events, or about the consequences of a decision, etc. Events are not simply “certain” or “uncertain”: we are capable of distinguishing degrees of probability.

Below we present a list of eight expressions, each denoting a degree of certainty. We ask you to order the expressions, that is, the expressions themselves, not the credibility of the sentence! Order the expressions by providing a number for each one of them. Start with a 1 for the highest degree of certainty and a subsequent number for each expression with a lower degree of certainty. If you find that some expressions denote the same level of certainty, you can indicate this by annotating them with the same number.

- It is **probable** that someone who must urinate often has a bladder infection.
 - It is **improbable** that someone with inflamed tonsils has no sore throat.
 - It is **possible** that someone faints from the heat.
 - It is **undecided** whether a crick of the back is caused by cramps or is related to a “slipped disc”.
 - It is **impossible** to prevent influenza.
 - It is **uncertain** whether or not a broken bone will heal completely.
 - It is **expected** that a stressed person will get a gastric ulcer.
 - It is **certain** that young people do not get varicose veins.
-

The following questionnaire was used in the third study of Chapter 6. The subjects were given a page with instructions, a page with four trial runs, and a booklet with all 28 possible pairs of expressions, each on a separate page. We will only present the instructions and one booklet page as an example.

Pairwise Comparisons of Verbal Probability Expressions

In this experiment you are asked to judge the comparability of a number of verbal probability expressions. Each form you have received contains two expressions. You are asked to compare these two expressions and indicate how similar or different you think these expressions are in meaning.

On each form you will find a line with the words *exactly the same* on one end and *completely different* on the other end. If, in your opinion, there is no difference in meaning between the two expressions, then place a mark at the beginning of the line at *exactly the same*. If you think there is a difference in meaning, place a mark somewhere on the line, indicating how much you think the expressions differ in meaning.

We would like to remind you that different people judge things differently. Two expressions that are similar to one person can be quite different to another person. This means that there are no right or wrong answers.

You will find one form for each pair of expressions. All you have to do is mark the line to indicate how similar or dissimilar you think a pair is. Wait about 5 seconds before proceeding to the next form and do not look back at the forms you have already filled in. The 32 pairs of expressions will require about ten minutes of your time.

To get used to the scoring procedure you will first be presented *four* trial pairs. After these, the real experiment starts.

possible - uncertain

exactly the same

completely different



The following questionnaire was used in the fourth study of Chapter 6. Each questionnaire consists of a instruction page, followed by six decision situations (A-F), each accompanied by a table similar to the one given in the instructions. We will present the instruction page and list the decision situations without the accompanying tables. For the decision situations listed, A, B and C are each presented with a table with numerical probability expressions. For situation A the expressions are ordered as in the table in the instructions, for situations B and C the order is 25, 85, 100, 15, 0, 75, 50, and 75, 25, 85, 50, 0, 100, 15, respectively. The situations D, E, and F are presented with a table with verbal probability expressions. The order of verbal expressions equals the order of their numerical translations used for the first three situations. Half of the subjects were given this version of the questionnaire; for the other half of the subjects situations A, B and C are combined with the tables with verbal probability expressions and D, E and F with the numerical expressions.

Decision Making under Uncertainty

In this experiment you will be presented with a number of hypothetical situation sketches. For each situation sketch there are eight different numerical or verbal expressions denoting the probability of occurrence of a certain event. In each case you are asked to make a certain decision and indicate how confident you are that this decision is correct. We will give an example:

The probability that someone who doesn't get enough sleep at night will have concentration problems is about ... %. You have an exam in the morning but are tired of studying. Will you go out for a while, tonight?

... %	Decision:	Confidence in this decision:
100 %	yes/no	complete ——— no
85 %	yes/no	complete ——— no
75 %	yes/no	complete ——— no
50 %	yes/no	complete ——— no
25 %	yes/no	complete ——— no
15 %	yes/no	complete ——— no
0 %	yes/no	complete ——— no

Fill in each of the probabilities from the table on the dots in the sentence, and circle "yes" or "no" for each alternative to indicate whether or not you would go out. Then indicate how much confidence you have in that decision by placing a mark on the horizontal line somewhere in the range of complete – no. We would like to remind you that different people judge things differently. This means that there are no right or wrong answers. All you have to do is circle "yes" or "no" for each alternative and mark the line that represents your confidence in the decision. There are 6 situation sketches, each with 8 alternatives that will require, in total, about ten minutes to answer.

- A) The probability that the railway-strike will be over tomorrow is about ... %. Tomorrow, you have an important appointment in another city. Will you cancel the appointment?
- B) The probability that it will rain tomorrow is about ... %. You have planned a big garden party for tomorrow. Will you cancel the party?
- C) Ms. T. has a non-serious physical complaint, which does however need treatment. The probability that Ms. T. is allergic to the usually prescribed drug H. is about ... %. Alternative drugs for her complaint are available, but these are less effective. Do you prescribe drug H. ?
- D) It is ... that/whether your patient X has broken his ankle. No x-ray can be made until tomorrow. Will you immobilise the patient in a plaster cast right now?
- E) It is ... that/whether you have a cavity in one of your teeth. It is not bothering you. Will you visit a dentist?
- F) It is ... that/whether an average student passes an exam without practising exercises. Suppose you are an average student and you want to pass the exam, but the practise group is at an inconvenient time. Will you join the practise group?

The following evaluation form was given to each of our domain experts and used for evaluating the elicitation method described in Chapter 6; the evaluation is described in Chapter 7. With each question, we left room for remarks or explanation.

Questions about the elicitation process

Please indicate the most suitable answer. Sometimes more than one alternative is possible. For each question there is room for remarks or an explanation.

- How did you experience assessing probabilities?
 - easy
 - efficient
 - natural
 - artificial
 - difficult

- How did you experience assessing probabilities compared to previous experiences?
 - easier
 - no difference
 - harder

- Is the use of text fragments a good way to indicate which probability is to be assessed?
 - yes
 - no, the following method would be better: . . .
- Is the use of text fragments helpful in assessing the probability itself?
 - yes
 - makes no difference
- The order of the required probabilities is such that the ones that should add up to 100% are grouped together. Does this aid the assessment of the probabilities?
 - yes, it is necessary
 - yes, it does make it easier
 - it makes no difference
- When asked to provide a probability, do you think in words or in numbers?
 - merely in words
 - merely in numbers
 - sometimes in words, sometimes in numbers
- When assessing a probability, what determines whether you do this in words or in numbers?
 - how sure I am of the answer
 - how familiar I am with the situation described
 - how hard it is for me to assess the probability
 - known values from the literature
 - something else, namely . . .
 - question is not applicable
- When estimating a probability, do you first think about what it should be and then mark the scale, or do you directly mark the scale at what seems the right position and then check whether the meaning is as intended?
 - first think, then mark
 - first mark, then check
 - I do something completely different, namely . . .
- Which part of the scale do you use to determine the position of the mark?
 - merely the words
 - merely the numbers
 - always both
 - sometimes words, sometimes numbers, sometimes both
 - other, namely . . .

- What is your opinion on the *number of words* on the scale?
 - sufficient
 - too few
 - too much

- What is your opinion on the *specific words* that are on the scale?
 - these are the words I use
 - I never use the following words: ...
 - the following words are missing: ...

- What is your opinion on the *position of the words* on the scale?
 - the positions are fine
 - the position of the following words differ from what I would expect: ...

- What is your opinion on the *number of numbers* that are on the scale?
 - sufficient
 - too few
 - too much

- What is your opinion on the *specific numbers* that are on the scale?
 - these are fine as anchors
 - the following numbers are superfluous: ...
 - the following numbers are lacking: ...

- Does it matter whether the scale is depicted vertically or horizontally?
 - a vertical scale is better, because ...
 - a horizontal scale is better, because ...
 - there's no difference

Bibliography

- [1] B. Abramson. ARCO1: An application of belief networks to the oil market. In B.D. D'Ambrosio, P. Smets, and P.P. Bonissone, editors, *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pages 1 – 8. Morgan Kaufmann Publishers, San Francisco, 1991.
- [2] B. Abramson, J. Brown, A. Murphy, and R.L. Winkler. Hailfinder: A Bayesian system for forecasting severe weather. *International Journal of Forecasting*, 12:57 – 71, 1996.
- [3] J.L. Anderson. Embracing uncertainty: The interface of Bayesian statistics and cognitive psychology. *Conservation Ecology [online]*, 2, 1998. <http://www.consecol.org/Journal/>.
- [4] S. Andreassen, M. Woldbye, B. Falck, and S.K. Andersen. MUNIN. A causal probabilistic network for interpretation of electromyographic findings. In J. McDermott, editor, *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 366 – 372. Morgan Kaufmann Publishers, Los Altos, California, 1987.
- [5] G.B. Baecher. Judgemental probability in geotechnical risk assessment. Technical report, prepared for The Office of the Chief, U.S. Army Corps of Engineers. http://www.enc.umd.edu/~gbaecher/papers.d/judge_prob.d/judge_prob.html.
- [6] I.A. Beinlich, H.J. Suermondt, R.M. Chavez, and G.F. Cooper. The ALARM monitoring system: a case study with two probabilistic inference techniques for belief networks. In J. Hunter, J. Cookson, and J. Wyatt, editors, *Proceedings of the Second Conference on Artificial Intelligence in Medicine*, pages 247 – 256. Springer-Verlag, Berlin, 1989.
- [7] E.S. Berner, G.D. Webster, A.A. Shugerman, J.R. Jackson, J. Algina, A.L. Baker, E.V. Ball, C.G. Cobbs, V.W. Dennis, E.P. Frenkel, L.D. Hudson, E.L. Mancall, C.E. Rackley,

- and O.D. Taunton. Performance of four computer-based diagnostic systems. *The New England Journal of Medicine*, 330:1792 – 1796, 1994.
- [8] R. Beyth-Marom. How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting*, 1:257 – 269, 1982.
- [9] C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-specific independence in Bayesian networks. In E. Horvitz and F.V. Jensen, editors, *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pages 115 – 123. Morgan Kaufmann Publishers, San Francisco, 1996.
- [10] W. Brun and K.H. Teigen. Verbal probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes*, 41:390 – 404, 1988.
- [11] G.D. Bryant and G.R. Norman. Expressions of probability: Words and numbers. *The New England Journal of Medicine*, 302:411, 1980.
- [12] D.V. Budescu and T.S. Wallsten. Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*, 36:391 – 405, 1985.
- [13] D.V. Budescu, S. Weinberg, and T.S. Wallsten. Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, 14:281 – 294, 1988.
- [14] W. Buntine. A guide to the literature on learning graphical models. *IEEE Transactions on Knowledge and Data Engineering*, 8:195 – 210, 1996.
- [15] U. Chajewska, L. Getoor, J. Normann, and Y. Shahar. Utility elicitation as a classification problem. In G.F. Cooper and S. Moral, editors, *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 79 – 88. Morgan Kaufmann Publishers, San Francisco, 1998.
- [16] U. Chajewska and D. Koller. Utilities as random variables: Density estimation and structure discovery. In G. Boutilier and M. Goldszmidt, editors, *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 63 – 71. Morgan Kaufmann Publishers, San Francisco, 2000.
- [17] D.A. Clark. *Psychological Aspects of Uncertainty and their Implications for Artificial Intelligence*. PhD thesis, Department of Applied Psychology, University of Wales Institute of Science and Technology, 1988.
- [18] R.T. Clemen and R.L. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19:187 – 204, 1999.
- [19] L.D. Cohn, M. Schydlower, J. Foley, and R.L. Copeland. Adolescents' misinterpretation of health risk probability expressions. *Pediatrics*, 95:713 – 716, 1995.

- [20] R.M. Cooke. *Experts in Uncertainty. Opinion and Subjective Probability in Science*. Oxford University Press, 1991.
- [21] G.F. Cooper. A method for using belief networks as influence diagrams. In *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*, pages 55 – 63, University of Minnesota, St. Paul, Minnesota, 1988.
- [22] G.F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393 – 405, 1990.
- [23] V.M.H. Coupé. *Sensitivity Analysis of Decision-Theoretic Networks*. PhD thesis, Faculty of Medicine and Health Sciences, Erasmus University Rotterdam, 2000.
- [24] V.M.H. Coupé, N.B. Peek, J. Ottenkamp, and J.D.F. Habbema. Using sensitivity analysis for efficient quantification of a belief network. *Artificial Intelligence in Medicine*, 17:223 – 247, 1999.
- [25] V.M.H. Coupé and L.C. van der Gaag. Properties of sensitivity analysis of Bayesian belief networks. *Annals of Mathematics and Artificial Intelligence*, 2000. To appear.
- [26] V.M.H. Coupé, L.C. van der Gaag, and J.D.F. Habbema. Sensitivity analysis: An aid for belief-network quantification. *Knowledge Engineering Review*, 15:1 – 18, 2000.
- [27] P. Dague. Symbolic reasoning with relative orders of magnitude. In R. Bajcsy, editor, *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1509 – 1514. Morgan Kaufmann Publishers, San Mateo, California, 1993.
- [28] R.B. Darlington. *Regression and Linear Models*, chapter 19. McGraw-Hill, Singapore, 1990.
- [29] A. Darwiche and M. Goldszmidt. On the relation between kappa calculus and probabilistic reasoning. In R. Lopez de Mantaras and D. Poole, editors, *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pages 145 – 153. Morgan Kaufmann Publishers, San Francisco, 1994.
- [30] F.J. Díez, J. Mira, E. Iturralde, and Z. Zubillaga. DIAVAL, a Bayesian expert system for echocardiography. *Artificial Intelligence in Medicine*, 10:59 – 73, 1997.
- [31] M.J. Druzdzel. *Probabilistic Reasoning in Decision Support Systems: From Computation to Common Sense*. PhD thesis, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1993.
- [32] M.J. Druzdzel. Qualitative verbal explanations in Bayesian belief networks. *Artificial Intelligence and Simulation of Behaviour Quarterly, special issue on Bayesian belief networks*, 94:43 – 54, 1996.

- [33] M.J. Druzdzel and M. Henrion. Efficient reasoning in qualitative probabilistic networks. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*. AAAI Press, Menlo Park, California, 1993.
- [34] M.J. Druzdzel and M. Henrion. Intercausal reasoning with uninstantiated ancestor nodes. In D. Heckerman and A. Mamdani, editors, *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 317 – 325. Morgan Kaufmann Publishers, San Mateo, California, 1993.
- [35] M.J. Druzdzel and H.J. Suermondt. Relevance in probabilistic models: “Backyards” in a “small world”. In *Working Notes of the AAAI 1994 Fall Symposium Series: Relevance*, pages 60 – 63, AAAI Press Technical Report FS-94-02. AAAI Press, California, 1994.
- [36] M.J. Druzdzel and L.C. van der Gaag. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. In Ph. Besnard and S. Hanks, editors, *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 141 – 148. Morgan Kaufmann Publishers, San Francisco, 1995.
- [37] M.J. Druzdzel and L.C. van der Gaag. “Building probabilistic networks: Where do the numbers come from?” — Guest editors’ introduction. *IEEE Transactions on Knowledge and Data Engineering*, 12:481 – 486, 2000.
- [38] W. Edwards. Unfinished tasks: A research agenda for behavioral decision theory. In R.M. Hogarth, editor, *Insights in Decision Making*, pages 44 – 65. The University of Chicago Press, Chicago, 1990.
- [39] W. Edwards. How to estimate thousands of reliable, valid probabilities. In M.J. Druzdzel, L.C. van der Gaag, M. Henrion, and F.V. Jensen, editors, *Working Notes of the IJCAI Workshop on Building Probabilistic Networks: Where Do the Numbers Come From?* AAAI Press, 1995. <http://www.sis.pit.edu/~dsl/hailfinder/probms2.html>.
- [40] W. Edwards. Hailfinder. Tools for and experiences with Bayesian normative modeling. *American Psychologist*, 53:416 – 428, 1998.
- [41] I. Erev and B.L. Cohen. Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes*, 45:1 – 18, 1990.
- [42] J.St.B.T. Evans and D.E. Over. *Rationality and Reasoning*. Psychology Press, Hove, United Kingdom, 1996.
- [43] N. Fenton. Probability elicitation and bias. <http://www.csr.city.ac.uk/people/norman.fenton/bbns/frconten.html>.
- [44] K. Fischer and H. Jungermann. Rarely occurring headaches and rarely occurring blindness: Is rarely = rarely? The meaning of verbal frequentistic labels in specific medical contexts. *Journal of Behavioral Decision Making*, 9:153 – 172, 1996.

- [45] A. Gifi. *Nonlinear Multivariate Analysis*. Wiley, Chichester, 1990.
- [46] G. Gigerenzer and U. Hoffrage. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102:684 – 704, 1995.
- [47] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84:57 – 112, 1996.
- [48] M. Gonzalez and C. Frenc-Mestre. Determinants of numerical versus verbal probabilities. *Acta Psychologica*, 83:33 – 51, 1993.
- [49] R. Greiner. Deployed Bayesian-nets systems in routine use. <http://www.cs.ualberta.ca/~greiner/BN-Routine.html>.
- [50] J.L. Gross and J. Yellen. *Graph Theory and its Applications*. The CRC Press series on discrete mathematics and its applications. CRC Press, Boca Raton, Florida, 1999.
- [51] Y. Ha and P. Haddawy. Problem-focused incremental elicitation of multi-attribute utility models. In D. Geiger and P.P. Shenoy, editors, *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 215 – 222. Morgan Kaufmann Publishers, San Francisco, 1997.
- [52] R.M. Hamm. Selection of verbal probabilities: A solution for some problems of verbal probability expression. *Organizational Behavior and Human Decision Processes*, 48:193 – 223, 1991.
- [53] F. Harary. *Graph Theory*. Addison-Wesley, Reading, Massachusetts, 1969.
- [54] D.E. Heckerman and B.N. Nathwani. Toward normative expert systems. Part I: The Pathfinder project. *Methods of Information in Medicine*, 31:90 – 105, 1992.
- [55] M. Henrion. Some practical issues in constructing belief networks. In L.N. Kanal, T.S. Levitt, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 161 – 173. Elsevier Science Publishers, North-Holland, 1989.
- [56] M. Henrion and M.J. Druzdzel. Qualitative propagation and scenario-based approaches to explanation in probabilistic reasoning. In P.P. Bonissone, M. Henrion, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 17 – 32. Elsevier Science Publishers, North-Holland, 1991.
- [57] M. Henrion, M. Pradhan, B. Del Favero, K. Huang, G. Provan, and P. O’Rorke. Why is diagnosis using belief networks insensitive to imprecision in probabilities? In E. Horvitz and F.V. Jensen, editors, *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pages 307 – 314. Morgan Kaufmann Publishers, San Francisco, 1996.
- [58] R.A. Howard and J. Matheson. Influence diagrams. In R.A. Howard and J. Matheson, editors, *Readings on the Principles and Applications of Decision Analysis*, pages 719 – 762. Strategic Decisions Group, Menlo Park, 1981.

- [59] D.C. Howell. *Statistical Methods for Psychology*. Wadsworth, Belmont, California, third edition, 1992.
- [60] E.K.R.E. Huizingh and H.C.J. Vrolijk. A comparison of verbal and numerical judgments in the Analytic Hierarchy Process. *Organizational Behavior and Human Decision Processes*, 70:237 – 247, 1997.
- [61] A.L. Jensen. Quantification experience of a DSS for mildew management in winter wheat. In M.J. Druzdel, L.C. van der Gaag, M. Henrion, and F.V. Jensen, editors, *Working Notes of the IJCAI Workshop on Building Probabilistic Networks: Where Do the Numbers Come From?*, pages 23 – 31. AAAI Press, 1995.
- [62] F.V. Jensen, U. Kjærulff, K.G. Olesen, and J. Pederson. Et forprojekt til et ekspertsystem for drift af spildevandsrensning (An expert system for control of waste water treatment – A pilot project). Technical report, Judex Datasystemer A/S, Aalborg, Denmark, 1989. In Danish.
- [63] J.B. Kadane and R.L. Winkler. Separating probability elicitation from utilities. *Journal of the American Statistical Association*, 83:357 – 363, 1988.
- [64] D. Kahneman, P. Slovic, and A. Tversky. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, 1982.
- [65] R.L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley & Sons, New York, 1976.
- [66] R.M. Kenney. Between never and always. *The New England Journal of Medicine*, 305:1097 – 1098, 1981.
- [67] A. Kong, G.O. Barnett, F. Mosteller, and C. Youtz. How medical professionals evaluate expressions of probability. *The New England Journal of Medicine*, 315:740 – 744, 1986.
- [68] M. Korver and P.J.F. Lucas. Converting a rule-based expert system into a belief network. *Medical Informatics*, 18:219 – 241, 1993.
- [69] P.J. Krause, S. Ambler, M. Elvang-Gøransson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11:113 – 131, 1995.
- [70] K. Kristensen and I. Rasmussen. A decision support system for mechanical weed control in malting barley. In H. Kure, I. Thysen, and A.R. Kristensen, editors, *Proceedings of the First European Conference on Information Technology in Agriculture*, pages 447 – 452. The Royal Veterinary and Agricultural University, Copenhagen, Denmark, 1997.
- [71] B. Kuipers, A.J. Moskowitz, and J.P. Kassirer. Critical decisions under uncertainty: Representation and structure. *Cognitive Science*, 12:177 – 210, 1988.
- [72] S.L. Lauritzen. *Graphical Models*. Oxford Statistical Science Series, vol. 17. Clarendon Press, Oxford, 1996.

- [73] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, 50:157 – 224, 1988.
- [74] A.V. Law, D.S. Pathak, and M.R. McCord. Health status utility assessment by Standard Gamble: A comparison of the probability equivalence and the lottery equivalence approaches. *Pharmaceutical Research*, 15:105 – 109, 1998.
- [75] Y. Lin and M.J. Druzdzel. Computational advantages of relevance reasoning in Bayesian belief networks. In D. Geiger and P.P. Shenoy, editors, *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 342 – 350. Morgan Kaufmann Publishers, San Francisco, 1997.
- [76] C.-L. Liu and M.P. Wellman. Incremental tradeoff resolution in qualitative probabilistic networks. In G.F. Cooper and S. Moral, editors, *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 338 – 345. Morgan Kaufmann Publishers, San Francisco, 1998.
- [77] C.-L. Liu and M.P. Wellman. Using qualitative relationships for bounding probability distributions. In G.F. Cooper and S. Moral, editors, *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 346 – 353. Morgan Kaufmann Publishers, San Francisco, 1998.
- [78] P.J.F. Lucas, H. Boot, and B.G. Taal. Computer-based decision-support in the management of primary gastric non-Hodgkin lymphoma. *Methods of Information in Medicine*, 37:206 – 219, 1998.
- [79] S.M. Mahoney and K.B. Laskey. Network engineering for complex belief networks. In E. Horvitz and F.V. Jensen, editors, *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pages 389 – 396. Morgan Kaufmann Publishers, San Francisco, 1996.
- [80] M.W. Merkhofer. Quantifying judgmental uncertainty: Methodology, experiences, and insights. *IEEE Transactions on Systems, Man, and Cybernetics*, 17:741 – 752, 1987.
- [81] J.F. Merz, M.J. Druzdzel, and D.J. Mazur. Verbal expressions of probability in informed consent litigation. *Medical Decision Making*, 11:273 – 281, 1991.
- [82] G.A. Miller. The magical number seven plus or minus two: Some limits on our capacity to process information. *Psychological Review*, 63:81 – 87, 1956.
- [83] M.G. Morgan and M. Henrion. *Uncertainty, a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, 1990.
- [84] F. Mosteller and C. Youtz. Quantifying probabilistic expressions. *Statistical Science*, 5:2 – 12, 1990.

- [85] M.A. Nakao and S. Axelrod. Numbers are better than words. Verbal specifications of frequency have no place in medicine. *The American Journal of Medicine*, 74:1061 – 1065, 1983.
- [86] B.J. O'Brien. Words or numbers? The evaluation of probability expressions in general practice. *Journal of the Royal College of General Practitioners*, 39:98 – 100, 1989.
- [87] M.J. Olson and D.V. Budescu. Patterns of preference for numerical and verbal probabilities. *Journal of Behavioral Decision Making*, 10:117 – 131, 1997.
- [88] S. Parsons. Refining reasoning in qualitative probabilistic networks. In Ph. Besnard and S. Hanks, editors, *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 427 – 434. Morgan Kaufmann Publishers, San Francisco, 1995.
- [89] S. Parsons. A proof theoretic approach to qualitative probabilistic reasoning. *International Journal of Approximate Reasoning*, 19:265 – 297, 1998.
- [90] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Palo Alto, 1988.
- [91] N.B. Peek and J. Ottenkamp. Developing a decision-theoretic network for a congenital heart disease. In E. Keravnou, C. Garbay, R. Baud, and J. Wyatt, editors, *Proceedings of the Sixth Conference on Artificial Intelligence in Medicine Europe*, pages 157 – 168. Springer Verlag, Berlin, 1997.
- [92] M. Pradhan, M. Henrion, G. Provan, B. Del Favero, and K. Huang. The sensitivity of belief networks to imprecise probabilities: An experimental investigation. *Artificial Intelligence*, 85:363 – 397, 1996.
- [93] O. Raiman. Order of magnitude reasoning. *Artificial Intelligence*, 51:11 – 38, 1986.
- [94] A. Rapoport, T.S. Wallsten, I. Erev, and B.L. Cohen. Revision of opinion with verbally and numerically expressed uncertainties. *Acta Psychologica*, 74:61 – 79, 1990.
- [95] L.K. Rasmussen. *Bayesian Network for Blood Typing and Parentage Verification of Cattle*. PhD thesis, Dina Research Report no. 38, Dina Foulum, Research Center Foulum, 1995.
- [96] E. Reiss. In quest of certainty. *The American Journal of Medicine*, 77:969 – 971, 1984.
- [97] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81 – 132, 1980.
- [98] S. Renooij. Probability elicitation for belief networks: Issues to consider. Submitted for publication.
- [99] S. Renooij. Qualitative probabilistic networks: the binary case. Master's thesis, Department of Computer Science, Utrecht University, 1996.
- [100] S. Renooij, S. Parsons, and L.C. van der Gaag. Context-specific sign-propagation in QPNs. Submitted for publication.

- [101] S. Renooij and L.C. van der Gaag. Decision making in qualitative influence diagrams. In D.J. Cook, editor, *Proceedings of the Eleventh International Florida Artificial Intelligence Research Society Conference*, pages 410 – 414. AAAI Press, Menlo Park, California, 1998.
- [102] S. Renooij and L.C. van der Gaag. Enhancing QPNs for trade-off resolution. In K.B. Laskey and H. Prade, editors, *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 559 – 566. Morgan Kaufmann Publishers, San Francisco, 1999.
- [103] S. Renooij and L.C. van der Gaag. Exploiting non-monotonic influences in qualitative belief networks. In *Proceedings of the Eighth International Conference on Information Processing and Management of Uncertainty*, pages 1285 – 1290, Madrid, Spain, 2000.
- [104] S. Renooij, L.C. van der Gaag, and S.D. Parsons. Propagation of multiple observations in QPNs. In A. van den Bosch and H. Weigand, editors, *Proceedings of the Twelfth Belgium-Netherlands Conference on Artificial Intelligence*, pages 235 – 242, Tilburg University, Tilburg, 2000.
- [105] S. Renooij, L.C. van der Gaag, S.D. Parsons, and S.D. Green. Pivotal pruning of trade-offs in QPNs. In G. Boutilier and M. Goldszmidt, editors, *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 515 – 522. Morgan Kaufmann Publishers, San Francisco, 2000.
- [106] S. Renooij and C.L.M. Witteman. Talking probabilities: Communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning*, 22:169 – 194, 1999.
- [107] J.A. Rice. *Mathematical Statistics & Data Analysis*. Duxbury Press, California, 1995.
- [108] M.P. Rutten-van Molken, C.H. Bakker, E.K. van Doorslaer, and S. van der Linden. Methodological issues of patient utility measurement. Experience from two clinical trials. *Medical Care*, 33:922 – 937, 1995.
- [109] T.L. Saati. *The Analytic Hierarchy Process*. McGraw-Hill, Inc., 1980.
- [110] L.J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.
- [111] R.D. Shachter. Evaluating influence diagrams. *Operations Research*, 34:871 – 882, 1986.
- [112] R.D. Shachter. Bayes-Ball: The rational pastime (for determining irrelevance and requisite information in belief networks and influence diagrams). In G.F. Cooper and S. Moral, editors, *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 480 – 487. Morgan Kaufmann Publishers, San Francisco, 1998.
- [113] M.A. Shwe, B. Middleton, D.E. Heckerman, M. Henrion, E.J. Horvitz, H.P. Lehmann, and G.F. Cooper. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR

- knowledge base I: the probability model and inference algorithms. *Methods of Information in Medicine*, 30:241 – 255, 1991.
- [114] D.C. Skinner. *Introduction to Decision Analysis*. Probabilistic Publishing, Gainesville, Florida, 1999.
- [115] K. Stavem. Quality of life in epilepsy: Comparison of four preference measures. *Epilepsy Research*, 29:201 – 209, 1998.
- [116] S.E. Stheeman, P.A. Mileman, M.A. van 't Hof, and P.F. van der Stelt. Blind chance? An investigation into the perceived probabilities of phrases used in oral radiology for expressing chance. *Dentomaxillofacial Radiology*, 22:135 – 139, 1993.
- [117] D.N. Stone and D.A. Schkade. Numeric and linguistic information representation in multiattribute choice. *Organizational Behavior and Human Decision Processes*, 49:42 – 59, 1991.
- [118] H.J. Suermondt. *Explanation in Bayesian Belief Networks*. PhD thesis, Report No. STAN-CS-92-1417, Departments of Computer Science and Medicine, Stanford University, Stanford, California, 1992.
- [119] Y. Takane, F.W. Young, and J. DeLeeuw. Nonmetric individual differences multidimensional scaling: An Alternating Least Squares method with optimal scaling features. *Psychometrika*, 42:7 – 67, 1977.
- [120] M. Tavana, D.T. Kennedy, and B. Mohebbi. An applied study using the Analytic Hierarchy Process to translate common verbal phrases to numerical probabilities. *Journal of Behavioral Decision Making*, 10:133 – 150, 1997.
- [121] D.R.M. Timmermans and P.A. Mileman. Lost for words: Using verbal terms to express probabilities in oral radiology. *Dentomaxillofacial Radiology*, 22:171 – 172, 1993.
- [122] P.A. Ubel, G. Loewenstein, D. Scanlon, and M. Kamlet. Individual utilities are inconsistent with rationing choices: A partial explanation of why Oregon's cost-effectiveness list failed. *Medical Decision Making*, 16:108 – 116, 1996.
- [123] L.C. van der Gaag and J.-J.Ch. Meyer. The dynamics of probabilistic structural relevance. In J.-J. Ch. Meyer and L.C. van der Gaag, editors, *Proceedings of the Eighth Dutch Conference on Artificial Intelligence*, pages 145 – 156, Utrecht University, Utrecht, 1996.
- [124] L.C. van der Gaag and J.-J.Ch. Meyer. Informational independence: Models and normal forms. *International Journal of Intelligent Systems*, 13:83 – 109, 1998.
- [125] L.C. van der Gaag, S. Renooij, B.M.P. Aleman, and B.G. Taal. Evaluation of a probabilistic model for staging of oesophageal carcinoma. In A. Hasman, B. Blobel, J. Dudeck, R. Engelbrecht, G. Gell, and H.-U. Prokosch, editors, *Medical Infobahn Europe: Proceedings of MIE2000 and GMDS2000*, pages 772 – 776. IOS Press, Amsterdam, 2000.

- [126] L.C. van der Gaag, S. Renooij, C.L.M. Witteman, B.M.P. Aleman, and B.G. Taal. How to elicit many probabilities. In K.B. Laskey and H. Prade, editors, *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 647 – 654. Morgan Kaufmann Publishers, San Francisco, 1999.
- [127] J. van Lenthe. ELI: An interactive elicitation technique for subjective probability distributions. *Organizational Behavior and Human Decision Processes*, 55:379 – 413, 1993.
- [128] J. Von Neumann and D. Morgenstern. *The Theory of Games and Economic Behavior*. Wiley, New York, third edition, 1953.
- [129] D. von Winterfeldt and W. Edwards. *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge, 1986.
- [130] T.S. Wallsten, D.V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth. Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, 115:348 – 365, 1986.
- [131] T.S. Wallsten, D.V. Budescu, and R. Zwick. Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science*, 39:176 – 190, 1993.
- [132] T.S. Wallsten, D.V. Budescu, R. Zwick, and S.M. Kemp. Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society*, 31:135 – 138, 1993.
- [133] E.U. Weber and D.J. Hilton. Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance*, 16:781 – 789, 1990.
- [134] M.P. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44:257 – 303, 1990.
- [135] M.P. Wellman and M. Henrion. Explaining “explaining away”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:287 – 291, 1993.
- [136] G. Wright and F. Bolger. *Expertise and Decision Support*. Plenum Press, New York, 1992.
- [137] N.L. Zhang and D. Poole. On the role of context-specific independence in probabilistic inference. In T. Dean, editor, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 1288 – 1293. Morgan Kaufmann Publishers, San Francisco, 1999.
- [138] A.C. Zimmer. Verbal versus numerical processing of subjective probabilities. In R.W. Scholz, editor, *Decision Making under Uncertainty*, pages 159 – 182. North-Holland, Amsterdam, 1983.

Samenvatting

Probabilistische netwerken worden veelvuldig toegepast voor het modelleren van onzekere informatie, en voor het redeneren met die onzekere informatie. Een probabilistisch netwerk is opgebouwd uit een kwalitatief deel en een kwantitatief deel. Het kwalitatieve deel bestaat uit een gerichte graaf waarin de knopen de belangrijke variabelen in het domein modelleren en de pijlen de afhankelijkheden tussen de variabelen aangeven. Voor elke variabele gerepresenteerd in de graaf zijn conditionele kansverdelingen gespecificeerd, die de kans vastleggen op een waarde van de variabele gegeven de verschillende waarde combinaties van zijn ouders in de graaf; deze verzameling van kansverdelingen vormt het kwantitatieve deel van een probabilistisch netwerk.

Het construeren van een probabilistisch netwerk begint gewoonlijk met de graaf: de belangrijke domein variabelen en hun waarden, tezamen met de afhankelijkheden tussen de variabelen, worden geïdentificeerd. Hiervoor wordt meestal de kennis van domein experts gebruikt. Alhoewel het construeren van een graaf veel tijd kost, is het toch een taak die als doenlijk beschouwd wordt. De graaf structuur van een probabilistisch netwerk legt vast welke kansverdelingen er geschat moeten worden. Voordat het netwerk gekwantificeerd wordt is het daarom van belang dat er een grote mate van zekerheid bestaat over de correctheid van de structuur. In dit proefschrift stellen we voor om de correctheid van de structuur te testen door er op een kwalitatieve manier mee te redeneren: alle pijlen in de graaf kunnen worden voorzien van een + of –-teken dat aangeeft of de ene variabele een positief of een negatief effect heeft op een andere variabele; deze tekens kunnen vervolgens door het netwerk gepropageerd worden. Op deze wijze kunnen we op een kwalitatieve manier de invloed van iedere willekeurige variabele op iedere andere willekeurige variabele bekijken; tevens kunnen de effecten van veranderingen in de structuur bestudeerd worden. Een ongekwantificeerd probabilistisch netwerk dat is voorzien van tekens op de pijlen heet een *kwalitatief probabilistisch netwerk*. In een kwalitatief probabilistisch netwerk zijn de tekens formeel gedefiniëerd in termen van de nog onbekende kansverdeling over de variabelen in de graaf; daarmee leggen de tekens, gegeven door domein experts, een verzameling eisen op aan de nog te verkrijgen kansverdelingen.

Het bepalen van de kansverdelingen voor de variabelen gerepresenteerd in de graaf, ofwel het kwantificeren van de graaf, vormt echter vaak een groot probleem. In veel domeinen lijkt er een enorme hoeveelheid probabilistische informatie voorhanden te zijn, maar in het algemeen moet het grootste deel van de kansen in het netwerk geschat worden door domein experts. Voor het eliciteren van kansen bestaat er een scala aan kans-elicitering methoden. Deze methoden zijn echter vaak erg complex en tijdsintensief in gebruik en daardoor ongeschikt voor het eliciteren van de duizenden kansen voor een probabilistisch netwerk. In dit proefschrift ontwerpen we een kans-elicitering methode die het mogelijk maakt om in een hoog tempo een groot aantal kansen te schatten. Deze kansschattingen hoeven niet noodzakelijkerwijs nauwkeurig te zijn: er bestaan methoden om de gevoeligheid van het netwerk voor onnauwkeurigheden in de schattingen te bepalen; het tijdrovende proces van het schatten van nauwkeurige kansen kan beperkt worden tot die kansen waarvoor het netwerk erg gevoelig is.

In dit proefschrift stellen we twee kwalitatieve benaderingen voor voor het kwantificeren van probabilistische netwerken. De eerste benadering is het gebruik van kwalitatieve probabilistische netwerken om tot een robuuste netwerk structuur en een verzameling eisen aan de benodigde kansverdelingen te komen. Het formalisme van kwalitatieve probabilistische netwerken is een bestaand formalisme, maar heeft een dusdanig hoog abstractie niveau dat eenvoudig nietszeggende antwoorden ('?'-tekens) worden afgeleid tijdens het redeneren. We stellen daarom in het eerste deel van het proefschrift verschillende uitbreidingen van dit formalisme voor, die de uitdrukingskracht van de kwalitatieve netwerken vergroten. In het tweede deel van het proefschrift ontwikkelen we een nieuwe elicitering methode. We breiden een bestaande elicitering methode uit met een aantal kwalitatieve componenten, zodat het eenvoudiger wordt voor mensen om kansen te schatten en we daardoor in een hoog tempo een groot aantal kansen kunnen eliciteren. Deze elicitering methode is de tweede kwalitatieve benadering die we voorstellen te gebruiken bij het kwantificeren van probabilistische netwerken.

Het eerste deel van het proefschrift gaat over kwalitatieve probabilistische netwerken en uitbreidingen ervan. In Hoofdstuk 2 geven we een overzicht van de kansrekening en de graaf-theorie die nodig is voor het begrip van kwalitatieve probabilistische netwerken. Kwalitatieve probabilistische netwerken worden besproken in Hoofdstuk 3. Een kwalitatief probabilistisch netwerk associeert kwalitatieve invloeden met de pijlen in zijn graaf. Voor een pijl $A \rightarrow B$ tussen een knoop A en een knoop B geeft een positieve kwalitatieve invloed aan dat voor hogere waarden van de variabele behorende bij knoop A , hogere waarden van B waarschijnlijker worden, onafhankelijk van de waarden van andere variabelen. Kwalitatieve invloeden tussen twee variabelen zijn dus gedefiniëerd onafhankelijk van andere variabelen gerepresenteerd in de graaf; kwalitatieve invloeden kennen bovendien alleen een notie van richting.

In Hoofdstuk 4 stellen we een aantal uitbreiding van het formalisme van kwalitatieve netwerken voor. De eerste uitbreiding is het expliciet maken van invloeden die niet monotoon zijn. Een invloed van een knoop A op een knoop B is niet monotoon als hij, bijvoorbeeld, positief is voor de ene waarde van een andere knoop C en negatief voor een andere waarde van C . In kwalitatieve probabilistische netwerken kunnen zulke invloeden alleen met een '?' aangeduid worden. Door expliciet te specificeren dat een invloed niet monotoon is en aan te geven welke knopen hiervan de oorzaak zijn, is het soms mogelijk om de non-monotonie op te lossen. De tweede uitbreiding is het toevoegen van een notie van sterkte: we maken onderscheid tussen

sterke en zwakke invloeden. Als er twee ketens in de graaf te vinden zijn tussen een knoop A en een knoop B zodat langs de ene keten de invloed positief is en langs de andere keten negatief, dan is het teken van de netto invloed van A op B onbekend: '?'. Weten we echter, bijvoorbeeld, dat de positieve invloed een sterke invloed is en de negatieve een zwakke, dan moet het netto resultaat positief zijn. De derde uitbreiding is het toevoegen van een notie van context. Door het gebruik van contexten kunnen we het teken van een kwalitatieve invloed af laten hangen van de waarden van knopen die niet rechtstreeks bij de invloed betrokken zijn. Zo kunnen we voor non-monotone invloeden niet alleen aangeven welke knoop de non-monotonie veroorzaakt, maar ook voor welke waarden, of contexten, de invloed positief dan wel negatief is. Dit geeft ons tevens de mogelijkheid om context-specifieke onafhankelijkheden expliciet te maken. Door dit soort uitbreidingen is het mogelijk om meer kennis in de netwerken te verwerken, waardoor er minder '?'-tekens resulteren tijdens het redeneren. Voor het geval er toch nog '?'s ontstaan tijdens het redeneren, hebben we een algoritme ontworpen dat kan achterhalen waar deze '?'s ontstaan zijn en welke informatie we nodig hebben om ze op te lossen. Tevens geven we een algoritme dat niet alleen het effect van een enkele observatie op de variabelen in het netwerk kan berekenen, maar ook het netto effect van meer observaties tegelijk. Dit algoritme is zo ontworpen dat onnodige '?'s voorkomen worden.

Het tweede deel van het proefschrift gaat over het eliciteren van kansen van domein experts. In Hoofdstuk 5 geven we een overzicht van bestaande elicitering methoden en wat er zoal komt kijken bij het eliciteren van kansen van experts. Hoofdstuk 6 bespreekt het onderzoek dat we verricht hebben naar het gebruik van verbale kansuitdrukkingen zoals 'waarschijnlijk' en 'onmogelijk'. We hebben proefpersonen een lijst met verbale kansuitdrukkingen laten genereren. Vervolgens hebben we een andere groep proefpersonen de meest genoemde uitdrukking laten rang ordenen en hebben we een groep de 'afstand' tussen de verschillende uitdrukkingen laten aangeven. Met behulp van de resultaten van dit onderzoek hebben we een kansschaal ontwikkeld waarop zowel verbale als numerieke kansexpressies staan. In onze nieuwe elicitering methode combineren we deze schaal met een stuk tekst dat de te schatten kans beschrijft. Daarnaast zorgen we ervoor dat alle kansen die betrekking hebben op dezelfde voorwaardelijke kansverdeling op een of twee opeenvolgende blaadjes aan de experts voorgelegd worden. Op deze wijze is het makkelijk te controleren of de gespecificeerde kansen wel tot 100% optellen. De methode, bestaande uit de combinatie van deze elementen, is met succes gebruikt met twee domein experts van het Nederlands Kanker Instituut/Antoni van Leeuwenhoekhuis voor het schatten van de kansen voor een probabilistisch netwerk voor slokdarm kanker. Een evaluatie van het gebruik van de methode, de verkregen schattingen en het gedrag van het netwerk is beschreven in Hoofdstuk 7. De evaluatie laat zien dat de methode erg goed bevalt, gemiddeld 170 kansen per uur oplevert en tot grove initiële schattingen leidt die vrij goed zijn.

In Hoofdstuk 8 vatten we kort samen wat we bereikt hebben. We hebben het formalisme van kwalitatieve probabilistische netwerken dusdanig uitgebreid dat we vaak sterkere resultaten af kunnen leiden tijdens het redeneren. We laten aan de hand van kwalitatieve abstracties van het slokdarm kanker netwerk en het bekende ALARM netwerk zien dat het gebruik van noties van sterkte en context de uitdrukkingskracht van 'echte' netwerken beduidend vergroot. Door de voorgestelde uitbreidingen zijn we beter in staat de robuustheid van de graaf te testen voordat het netwerk gekwantificeerd wordt, en verkrijgen we een grotere verzameling eisen aan de nog

te schatten kansverdelingen. Voor het daadwerkelijk kwantificeren van probabilistische netwerken met behulp van domein experts hebben we een nieuwe kans-elicitering methode ontwikkeld. Deze methode maakt het experts mogelijk om een grote hoeveelheid initiële kansen in korte tijd te schatten. Hoofdstuk 8 geeft tevens wat richtlijnen voor verder onderzoek in de richting van kwalitatieve netwerken en kans-elicitering. Daarnaast is er nog veel onderzoek vereist naar uitleg van probabilistische netwerken; wellicht kan ons algoritme voor het achterhalen waar ‘?’s ontstaan daar aan bijdragen. Tenslotte gaan we kort in op beslissingsondersteuning met behulp van probabilistische netwerken. Als we de door ons beschouwde netwerken uitbreiden met ‘beslis-knoppen’, die mogelijke beslissingen modelleren, en een ‘utiliteitsknoop’, die de wenselijkheid van bepaalde beslissingen in een bepaalde context aangeeft, dan krijgen we een *influence diagram*. Voor een *influence diagram* moeten ook utiliteiten voor de utiliteitsknoop geschat worden; dit laatste is zo mogelijk nog moeilijker dan het schatten van kansen. Een kwalitatieve abstractie van een *influence diagram* is een mogelijke eerste stap in de richting van het kwantificeren van *influence diagrams*.

Dankwoord

Het einde van een proefschrift en promotie traject is een mooi moment om eens terug te kijken en iedereen die hierin een rol vervuld heeft de revue te laten passeren. Allereerst wil ik mijn dagelijkse begeleidsters Linda van der Gaag en Cilia Witteman bedanken voor de intensieve samenwerking. Onze bijeenkomsten en uitstapjes waren nuttig en ook erg gezellig en leuk: het is geen wonder dat we eens een *soar throat* aan het lachen overgehouden hebben. Linda, bedankt voor het enthousiasme dat je toonde bij elke stap die ik zette. Je nam de tijd om alles wat ik produceerde zeer uitgebreid van commentaar te voorzien en zo heb ik ook stukje bij beetje (grotendeels) afgeleerd om ‘straat taal’ te schrijven. Cilia, dankzij jou weet ik nu enigszins wegwijs in een wereld die voor mij totaal onbekend was: de experimentele psychologie. Samen hebben we heel wat gelezen, experimenten opgezet, proefpersonen lastig gevallen en data verwerkt.

Naast de enorme inzet van Linda en Cilia is er nog een aantal mensen dat heeft bijgedragen aan de huidige vorm en inhoud van het proefschrift. Als eerste wil ik Babs Taal en Berthe Aleman van het Nederlands Kanker Instituut/Antoni van Leeuwenhoekhuis heel erg bedanken dat ze het aangedurfd hebben om een hernieuwde poging te doen tot het schatten van de kansen voor het oesofagus netwerk. Dankzij alle tijd die zij bereid geweest zijn er vrijwillig in te steken, hebben wij onze elicitation methode kunnen testen en tevens het netwerk in een stadium gekregen waarin we het konden gaan evalueren. I am also very grateful to Shaw Green, Simon Parsons and Peter McBurney for the fruitful and fun meeting we had in London, and, with Shaw, in Utrecht. Simon, I very much enjoyed our encounters at several conferences, and am looking forward to visiting you again, some day soon, in Liverpool, and finally get a chance at tasting that sticky toffee pudding. Thank you, as well, for volunteering to be a member of my reading committee. For reviewing this thesis, I should also thank Arie de Bruin, Richard Gill, Karl Halvor Teigen, and Michael Wellman.

Voor hun advies met betrekking tot de precieze opzet van sommige experimenten en met name over welke statistische methoden los te laten op onze data en hoe de uitkomsten daarvan te interpreteren, ben ik dank verschuldigd aan Pieter Koele, Harald Kunst, en Herbert Hoijtink. Ook alle stafleden van het instituut en alle studenten Informatica, CKI, Geneeskunde en Medische

Biologie, die hebben deelgenomen aan onze experimenten, wil ik hierbij bedanken. Tevens dank ik John-Jules Meyer voor het becommentariëren van het proefschrift en Niels Peek voor zijn enthousiasme over en uitgebreide commentaar op Hoofdstuk 4 en 6. Ik wil Eveline Helsper bedanken voor het verzorgen van chocolaatjes bij de laatste loodjes: dat vond ik erg lief!

Naast werk was er gelukkig ook voldoende afleiding. Regelmatig kwamen er mensen binnen lopen voor een praatje met mij en/of mijn kamergenoten. Met mijn kamergenoten heb ik heel wat afgepraat, gespeeld en gepuzzeld en ik denk daarom met veel plezier terug aan de tijden dat ik een kamer deelde met Martijn Schrage, Arend-Jan de Groot, Daan Leijen en, weliswaar geen kamergenoot maar toch vaak aanwezig, Arjan van IJendoorn. Martijn, ik vond het erg prettig om een goede bekende als kamergenoot te hebben; bovendien vulden we elkaar goed aan bij het oplossen van U-blad en Sum crypto. Ik dank jullie allen voor het verzorgen van precies dat beetje afleiding dat je nodig hebt om hard te kunnen werken.

Tot slot wil ik Papa en Mama, en Jorinde en Martijn bedanken voor hun hulp, advies, steun en interesse de afgelopen jaren. Ook Jenny, Wijnand, Betsy, Richard, Marie-louise, Peter, Sacha en overige familie en vrienden wil ik hiervoor hartelijk bedanken. Papa, ik vind het geweldig dat je mijn hele proefschrift doorgeworsteld hebt en ik dank je voor je tekstuele commentaar. Lieve Edwin, ik wil jou tot slot heel erg bedanken voor: al je steun, hulp en advies; het meerdere malen optreden als proef-proefpersoon; je hulp bij het oplossen van een, achteraf gelukkig onnodige, vergelijking met 20 constraints in 6D op een warme zomer-namiddag aan het Henschotermeer; het prachtig ray-trace-en van de dobbelsteentjes op de kaft; en nog veel meer. . . kom op hè, nu jij nog!

Klaar.

Curriculum Vitae

Silja Renooij

13 oktober 1972

Geboren te Amersfoort.

september 1983 – juni 1990

Atheneum aan de Scholengemeenschap De Amersfoortse Berg in Amersfoort.
Diploma behaald in juni 1990.

september 1990 – augustus 1996

Studie Informatica aan de Vakgroep Informatica van de Universiteit Utrecht.
Doctoraal diploma behaald in augustus 1996.

januari 1997 – december 2000

Assistent in Opleiding bij het Instituut voor Informatica en Informatiekunde van de Universiteit Utrecht.

januari 2001 – heden

Post-doc bij het Instituut voor Informatica en Informatiekunde van de Universiteit Utrecht.

List of Notations

Graphs

G	directed graph	14
$V(G)$	set of nodes in G	14
$A(G)$	set of arcs in G	14
$V(t)$	set of nodes on trail t	15
$A(t)$	set of arcs on trail t	15
$\pi_G(A)$	parents of node A in G	14
$\pi_G^*(A)$	ancestors of node A in G	14
$\sigma_G(A)$	children of node A in G	14
$\sigma_G^*(A)$	descendants of node A in G	14
t^{-1}	inverse of trail t	15
$t_i \circ t_j$	concatenation of trails t_i and t_j	16
$t_i \parallel t_j$	parallel composition of trails t_i and t_j	16
$\langle X Y Z \rangle_G^d$	d-separation of X and Z by Y in G	16

Basic Qualitative Probabilistic Networks

Q	qualitative probabilistic network	37
Δ	set of all hyperarcs modelling qualitative properties	37
δ	sign of a qualitative property	
$\delta_i \otimes \delta_j$	sign-product of δ_i and δ_j	25
$\delta_i \oplus \delta_j$	sign-sum of δ_i and δ_j	25
$S^\delta(A, B)$	qualitative influence with sign δ of A on B along $A \rightarrow B$	21
$\hat{S}^\delta(A, B, t)$	qualitative influence with sign δ of A on B along trail t	23
$Y^\delta(\{A, B\}, C)$	additive synergy with sign δ of A and B on C for arcs	28
$\hat{Y}^\delta(\{A, B\}, C, \{t_i, t_j\})$	additive synergy of A and B on C for trails t_i and t_j	29

$X^\delta(\{A, B\}, c_0)$	product synergy of A and B w.r.t. c_0 for arcs	33
$\hat{X}^\delta(\{A, B\}, c_0, \{t_i, t_j\})$	product synergy of A and B w.r.t. c_0 for trails t_i, t_j	33
$Z^\delta(A, B, c_0)$	intercausal influence of A on B induced by c_0	34
$\hat{Z}^\delta(A, B, c_0, t)$	intercausal influence of A on B along t , induced by c_0	34

Refined Qualitative Probabilistic Networks

\sim_P	sign of non-monotonicity provoked by P	53
α	cut-off value for strong/weak influences	57
δ^i	sign augmented with multiplication index	60
$S^{\delta^1}(A, B)$	enhanced influence of A on B along arc $A \rightarrow B$	57
$\hat{S}^{\delta^i}(A, B, t)$	enhanced influence of A on B along trail t	60
$\delta^i \otimes \delta^j$	enhanced sign-product of δ^i and δ^j	61
$\delta^i \oplus \delta^j$	enhanced sign-sum of δ^i and δ^j	63
$\delta(X)$	context-specific sign for the set of nodes X	76
c_X	context for the set of nodes X	76
ϵ_X	empty context for the set of nodes X	76
\mathcal{C}_X	set of all contexts c_X for the set of nodes X	76
$X(O)$	exclusion set for node O	97

Miscellaneous

Pr	joint probability distribution	12
B	probabilistic network	18
α	significance level in hypothesis tests	177
χ^2	goodness-of-fit measure	181
SD	standard deviation	