DOCTORAATSPROEFSCHRIFT

2004

Faculteit
Toegepaste Economische Wetenschappen

Customer Satisfaction and Loyalty Research: A Bayesian Network Approach

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Toegepaste Economische Wetenschappen
te verdedigen door

Waldemar JAROŃSKI

Promotor : Prof. dr. J. Bloemer
Co-promotor : Prof. dr. K. Vanhoof

LIMBURGS UNIVERSITAIR CENTRUM
PARTNER IN DE UNIVERSITEIT LIMBURG

2004

Faculteit
Toegepaste Economische Wetenschappen

Customer Satisfaction and Loyalty Research:

A Bayesian Network Approach

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Toegepaste Economische Wetenschappen
te verdedigen door

Waldemar JAROŃSKI

Promotor : Prof. dr. J. Bloemer
Co-promotor : Prof. dr. K. Vanhoof

LIMBURGS
UNIVERSITAIR
CENTRUM
PARTNER IN DE UNIVERSITEIT LIMBURG

According to the guidelines of the Limburgs Universitair Centrum, a copy of
this publication has been filed in the Royal Library Albert I, Brussels,as publication
D/2004/2451/53

Acknowledgments

This work would not have been possible without support from many people. Let me take this opportunity to acknowledge their help.

First of all, I wish to express my deepest gratitude to my research supervisors, Prof. Dr José Bloemer from the University of Nijmegen and Prof. Dr Koen Vanhoof from the Limburgs Universitair Centrum for their scientific guidance, encouragement and their invaluable support in finalizing this thesis.

Furthermore, I am indebted greatly to the external reviewers of this work, Prof. Dr A. Kuijlen from Tilburg University and Prof. Dr E. Nijssen from the University of Nijmegen for their valuable remarks and critiques that significantly improved the final shape of this thesis. I would like to give my warmest thanks also to the jury members and colleagues from the LUC, Prof. Dr Geert Wets who encouraged me to do research in the exciting field of Bayesian networks, and Dr Tom Brijs for their useful comments and the many discussions from which this work benefited.

Z całego serca chciałbym podziękować moim najbliższym, mojej Mamie, bratu i całej rodzinie za wsparcie duchowe w trakcie mojej pracy nad tą rozprawą i za to, że dzięki nim jestem tym człowiekiem, którym jestem. Dziękuję Oldze za radość którą wniosła do mojego życia.

Ponadto, pragnę wyrazić wdzięczność gronu profesorskiemu oraz pracownikom naukowym Akademii Ekonomicznej we Wrocławiu, a zwłaszcza Prof. dr hab. E. Niedzielskiej oraz śp. Prof. dr. hab. A. Baborskiemu za umożliwienie mi podjęcia studiów doktoranckich, oraz koleżankom i kolegom z Katedry Systemów Sztucznej Inteligencji za ich moralne wsparcie podczas mojej pracy na AE.

I would also like to thank all my friends and colleagues at the LUC, with whom I have had the pleasure of working over the years, especially my officemates: Birgit, Elke, Enid, Ellen, Benoit, as well as Karolien, Fung, Davy, Filip, Isabel, Tamara, Geert, Kris, Evy, Liesbeth, Mark, Bram and many others. I would like to give special thanks to Prof. Dr Tomasz Burzykowski and Prof. Dr Ziv Shkedy for their invaluable help when I hassled them with my questions concerning statistics.

Last but not least, I also thank my great housemates from Leuven: Leen, Goedele, Sofie, Sanju, and Sam for the wonderful time I spent with them and for their understanding when I had to refuse to join them on many occasions. I will also have in my memory other good friends Aneta, Artur, Michał, Marek, Ewa, Hans and others who supported me and contributed to making my stay in Flanders an unforgettable experience.

Waldemar Jaroński
14 September 2004

Table of contents

1. Introduction

1.1. Background

Bayesian data analysis has recently gained substantial interest also in the marketing research community [Rossi and Allenby, 2003; Wedel *et al.*, 1999; Shively *et al.*, 2000]. Bayesian statistics, compared with traditional statistics, offers new opportunities for scientific investigation, in which new empirical data can be combined with historical data and accumulated knowledge for a holistic validation and justification of empirical laws and theories. The Bayesian approach enables theoretically and philosophically sound estimation of model parameters and interpretation of results. The approach proved especially valuable when the information sources are scarce, or when data come from sources of different kinds. Among the most recognized Bayesian data analysis techniques, we can perhaps identify hierarchical Bayes models, and Bayesian regression as those that find their successful application in social sciences. In marketing, these methods have also been applied in a wide range of problems from new product strategies to pricing [Rossi and Allenby, 2003].

One of other techniques that has its roots in Bayesian analysis are Bayesian networks. Mathematically, Bayesian networks, or BNs, are network-based framework for representing and analysing models involving uncertainty and allow for concise and effective representation of the joint probability distribution over random variables in a domain under consideration. Historically, the Bayesian network formalism was developed and known long ago in the statistics community, but due to serious computational bottlenecks they were out of use. Only recently, since they were popularised in the artificial intelligence community by Pearl's work [1988], new successful developments have been proposed enabling their use in a wide range of problems. Especially, in areas involving uncertainty, such as medical diagnosing [Andreassen *et al.*, 1991; Heckerman *et al.*, 1992], troubleshooting [Heckerman et al., 1995; Jensen et al., 2001], decision support, automatic speech recognition [Zweig and Russel, 1999], their application proved valuable.

The unique contribution of this work comes mainly from the intersection of the Bayesian network literature and the marketing modelling literature. In spite of their apparently attractive features for solving various marketing problems, Bayesian networks are, to the best of our knowledge, still not a well-recognized technique within the marketing community [Lilien and Rangaswamy, 2000, p.232]. This lack of recognition can be attributed to the following general reasons. Most importantly, we acknowledge that the methodology is still in the early stages of its maturity with respect to specific requirements of marketing research. In the context of causal modelling, some authors make even a parallel between the current stage of development of Bayesian networks with the stage of structural equation models in the 1970s [Anderson and Lenz, 2001]. Secondly, even taking into account its relative immaturity their use is appealing; however, there is a lack of a thorough discussion of basic features and potential added value of the Bayesian network technology as a tool in the arsenal a marketing researcher.

This thesis is also motivated with the observation that little attention has been paid to date on adapting or evaluating Bayesian networks as a potential technique for conducting research, let alone marketing research. Instead, since its bloom in the 1990's, the vast majority of research on Bayesian networks has been focused rather on developing algorithms and fostering technical innovations for the purpose of expert systems. As such, this previous work has been limited to problems existing in artificial intelligence and data mining.

Since we find it very important to bring the Bayesian network approach closer to marketing, as the overall goal of this thesis we aim to provide a critical evaluation of the application of Bayesian networks in theoretical and practical marketing research, and propose new methods and developments within the Bayesian network modelling to improve its current abilities with respect to specific requirements existing in the marketing research. However, this formulation of the overall objective would require an immense, if not unfeasible, task due to plethora of avenues in marketing research; therefore, we constrain ourselves to only one particular area in marketing science: the Customer Satisfaction and Loyalty (CS&L) research. Due to the growing importance of e-commerce and Internet in marketing science [e.g., Mahajan and Venkatesh, 2000; O'Connor and Galvin, 2001], we will consider the CS&L phenomenon both in the traditional, "mortar-and-brick" context as well as in the online one.

Furthermore, the critical evaluation that we undertake in this thesis should be regarded as internal validation rather than external one. In other words, it is our aim to examine Bayesian networks individually rather than to compare this methodology in a competitive setting with other techniques applied today in CS&L research in terms of their respective outcomes and findings and to establish which techniques are superior and which perform worse. Consequently, we take the position by which the Bayesian network approach is considered in this thesis merely as another approach that can help understand and research the CS&L phenomenon.

In order to achieve the aforementioned overall goal of the dissertation, it is also essential that the perspective we take here be from the position of a CS&L scientist rather than of a Bayesian network expert. In other words, we will take the needs and objectives in CS&L research as the starting point for this discussion. Consequently, let us now present our diagnosis of the requirements existing today in CS&L research and mark in more detail the areas in which the Bayesian network literature is still missing.

1.2. Research motivation

The formulation of our overall research goal requires that we delineate at this point between two streams of investigation within CS&L research: theoretical (also referred to as basic or academic), and practical (applied).

1.2.1. Theoretical versus practical CS&L research

The difference between theoretical and practical Customer Satisfaction & Loyalty research can be best explained by the known Three Dichotomies Model of the scope of marketing proposed by Philip Kotler and first published in an article by Shelby Hunt [Hunt, 1976]. The model proposes that all marketing phenomena, topics or issues can be categorised using the three dichotomies: 1) profit/non-profit sector, 2) micro/macro dimension, and 3) positive/normative research. In the context of this work, the most important dichotomy is the third one, as the work is clearly situated in the profit sector and the micro dimension. The positive marketing adopts the perspective of attempting to describe, explain predict and understand the marketing phenomena in focus; this perspective examines "what is". In contrast, the normative perspective attempts to prescribe what organisations ought to do, that is, it examines "what ought to do" [Hunt, 1991].

Consequently, we will adopt the view by which the theoretical CS&L research can be identified with the positive perspective on marketing, and the normative perspective corresponds with the practical CS research. Let us first define the focus of theoretical CS&L research.

Customer Satisfaction is accepted as a critical concept in marketing thought and consumer research [e.g., Yi, 1990; Peter and Olson, 1996; Erevelles and Leavitt, 1992]. Moreover, Customer Satisfaction is a concept that has very significant consequences for our society, since as a result, the overall quality of life is expected to be enhanced [Yi, 1990]. It is a critical asset for all organizations. For non-profit organizations the objective is to enhance trust and contentment of donators. For companies, it is to facilitate marketing decisions in companies in order to keep their customers satisfied and loyal. It is also important for national and local administration to track the satisfaction of citizens at a national level [Fornell, 1992; The International Foundation for Customer Focus, 2000]. In this work, we put emphasis predominantly on Customer Satisfaction in the private sector.

In spite of the crucial role that satisfaction plays in retaining customers [Rust and Zahorik, 1993], and higher future profitability [Anderson *et al.*, 1994], it can be viewed merely as a necessary step in loyalty formation [Oliver, 1999]. The ultimate goal for customer-oriented companies is customer loyalty as a proxy for profitability [Reichheld and Sasser, 1990; Fornell *et al.*, 1996]. However, the relation between customer satisfaction and loyalty is not well-specified and still remains to be investigated [Oliver, 1999], and hence the name of the field of research.

We can formally state that the aim of Customer Satisfaction and Loyalty research is identification of post-purchase cognitive, affective, and normative processes, through which consumers become satisfied, and eventually loyal, to a service/product provider. In other words, the focus of the theoretical CS&L research is to develop the theory of the CS&L phenomenon.

The focus of the practical CS&L research is quite different. It occurs that the main interest of marketing practitioners lies not so much in theoretically sound conceptual models of CS&L, but in models that more directly let them support their marketing decisions. Whereas in academic CS&L research the relation between customer satisfaction and loyalty is studied thoroughly, here this relation seems usually to be assumed true and as such is not so much in the focus of interest. The objective of practical CS&L research is furthermore not contributing to the marketing knowledge in general; in contrast, it aims to solve some specific problems that a company is facing. Typically, the focus of analysis goes to the link between the importance of service/product attributes for driving overall customer satisfaction with the service/product; this relationship can vary across companies and is much dependent on the unique features of the service/product and industry. Once the importance is derived from data, the next step is determining the performance of the attributes. The underlying goal of this stream of research is to accommodate the company resources in an optimal way, for instance, in terms of assigning more resources to those attributes that drive overall satisfaction, and require improved performance, and allocate less resources in these aspects of the service/product in which possible overkill exists. This analysis is often defined in the marketing literature as importance/performance analysis [e.g., Martilla, 1977], and we will refer to this stream in marketing research as practical Customer Satisfaction (CS) research in this thesis.

1.2.2. The requirements in theoretical CS&L research

In this section, we discuss the requirements of theoretical CS&L research in relation to Bayesian network analysis. Clearly, the aim of the theoretical CS&L research is to develop the theory of Customer Satisfaction and Loyalty phenomenon.[1] To make this consideration more precise, let us first define terms as theory, and model, and specify the relation between a theory and a model.

In this thesis, we will adopt the definition of theory according to which a theory is "a systematically related set of statements, including some lawlike generalizations, that is empirically testable" [Rudner, 1966; Hunt, 1991]. The purpose of theory is to increase scientific understanding through a systemized structure capable of both explaining and predicting phenomena [*idem*]. A less formal definition of a scientific theory states that it is "a system of ideas and

[1] In fact, the discussion here would be relevant for many other theories in marketing, and other social sciences in general.

observations, all related among themselves in a meaningful way" [e.g., Feigl, 1970].

A theory must specify how various constructs are interrelated, and this is the point where (marketing) models come into play. According to Lazer [1962], a model "is simply the perception or diagramming of a complex or a system" and is "the base for marketing theories, since they are the axioms or assumptions on which marketing theories are founded." Rigby suggests that a model is any structure that purports to represent something else [Rigby, 1965]. In this thesis we accept the position that a model is a formal representation of a theory. In the same sense, we will also apply the term "a theoretical model". Theoretical models attempt to precisely and parsimoniously characterise the world in which a phenomenon of interest can be shown to occur [Rangaswamy, 1993]. All models make a set of assumptions or suppositions that do not always correspond exactly with a real marketing environment in question. Usually, models are employed to simplify an existing marketing world. It follows that all theories are models but not all models are theories [Hunt, 1991].

The question here is how we can tell whether a specific structure can be viewed as a theory or not. In order to define the requirements of a theory, it would be at this point worthwhile to delineate first between the context of theory discovery and the context of theory justification in philosophy of science. Hunt [1991] argues on making a clear delineation of these two contexts in marketing science, as he suggests that only then it is possible for the marketing science to advance and produce long-lasting results and solid theories. The context of discovery is concerned with how one goes about discovering hypotheses, theories, and laws, or what kind of procedures, activities or rules will assist the researcher in uncovering them. In contrast, issues such as how one scientifically explains phenomena, or what are procedures or rules that delineate the criteria for accepting or rejecting knowledge belong in the context of justification (validation, corroboration, confirmation).

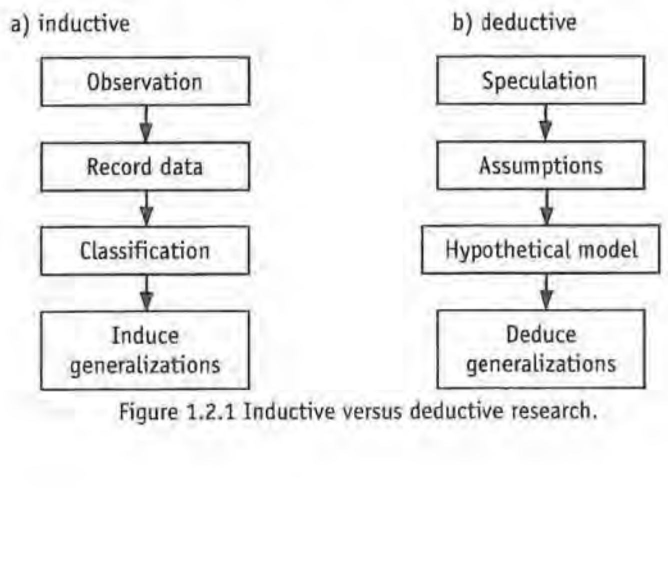


Figure 1.2.1 Inductive versus deductive research.

Let us first describe in a little more detail the context of discovery. In Figure 1.2.1, two well-known routes for scientific discovery are shown.[1] The first route in Fig. 1.2.1a) can be seen as a strict inductivist approach, whereas the route in Fig. 1.2.1b) is the deductivist route. The inductive research in the strict sense starts typically with making observations about the world and recording them as data; next, the data are rearranged and analysed so as to "bring order out of chaos"; lastly, lawlike generalizations or patterns are induced [McGarry, 1936]. In the deductivist approach, we start by making speculations about a theory, forming assumptions and advancing hypotheses; next, we proceed by proposing a hypothetical model, and ultimately, we can deduce generalizations [Hunt, 1991]. It is worth noting that in the marketing research literature, the inductive approach is sometimes also referred to as the exploratory approach [Armstrong *et al.*, 2001, p. 171].

Another important issues in developing theories, the CS&L theory including, that contribute to scientific understanding are the issues of moderating effects, and mediating (intervening) variables. These phenomena provide more explanatory power for relationships between concepts in a theory [e.g., Cooper and Emory, 1995]; similarly, Bagozzi [1989, 1994b] argues that marketing modelling techniques should possess the potential of accounting for such situations.

With regard to the context of justification, one must address the issue of how one delineates whether the model can be deemed explanatory or not. It is also in this context that one must consider what are the criteria of scientifically theories. In the literature on the philosophy of marketing science, we can find that the requirements of any theory, and therefore also of any theory referring to marketing phenomena in particular, can be classified as description, prediction, and explanation [Hunt, 1983, 1991; Rositer, 1994]. Likewise, a theory of Customer Satisfaction & Loyalty must also achieve these three goals. According to Hunt [1991], there can four main normative criteria be specified to decide whether or not accept the model as explanatory: 1) first of all, it must show that the phenomenon to be explained was somehow expected to occur, 2) be intersubjectively certifiable, 3) have empirical contents and be empirically testable, and 4) be pragmatic.

In order for a theory to be empirically testable, we must be able to form concepts and hypotheses, and make observations and measurements [Kaplan, 1964]. In the social sciences, and so in the CS&L research, psychological constructs are treated as latent concepts, which cannot be measured directly. Instead, multiple-item measurement instruments are necessary to capture the entire character of the construct indirectly [e.g., Bagozzi, 1994a]. In the marketing modelling literature, this area is known as the theory of measurement

[1] In the context of discovery, probably the most successful discoveries have been accomplished in line with the "Eureka" route, as the flash of perceptual insight.

[*idem*]. In short, it involves defining latent constructs, proposing empirical definitions, and determining internal consistency and validity of constructs.

It must be noted that, to the best of our knowledge, Bayesian networks have not been addressed or discussed to date with respect to these above mentioned requirements in the theoretical CS&L research, i.e., concerning inductive and deductive theory development, issues in the context of scientific justification, measurement modelling issues, or moderating and mediating effects, in the literature including both the Bayesian network literature and the marketing literature. To be more specific, one might argue that the issue of hidden node models, which is an actively researched area within the Bayesian network community [see for instance: Friedman, 1998; Heckerman *et al.*, 1999; Rusakov and Geiger, 2003] could be seen as an example of latent construct modelling; however, in these and other articles in the BN literature, hidden nodes are defined in the sense of a variable omitted from the model; Furthermore, no attention has been paid to latent construct modelling, and measurement models in the sense of the theory of measurement. Similarly, none of the articles that have appeared within the marketing literature addresses the above-mentioned requirements. We can therefore conclude that there exists a gap in the marketing literature that we aim to fill.

1.2.3. The use of Bayesian networks in practical CS studies

One of the primary tasks in practical customer satisfaction studies carried out by companies and other organisations pertains to determining product/service factors driving satisfaction and/or dissatisfaction [Oliver, 1997; Hill and Alexander, 2000]. The managerial results of such a study should identify possible factors as priorities for improvement to focus company resources on these factors that require better performance on the one hand, and to decrease resources on those that possibly do not have a link with satisfaction on the other hand [Naumann and Giel, 1995]. In other words, the findings of such a study should provide insight as to the importance of product/service dimensions in terms of the strength of their influence on overall (dis)satisfaction and the character of this influence.

However, findings suggest that the relationships between performance of product/service features and overall satisfaction can often be non-linear and not straightforward. For example, Mittal *et al.* [1998] investigated this link and found that attribute-level performance impacts satisfaction differently based on whether consumer expectations were positively or negatively disconfirmed. In their study, overall satisfaction was found to be sensitive to changes in low attribute levels, whereas at high levels of attribute performance, overall satisfaction showed diminished sensitivity.

In light of these findings of a complex nature of the relationships between attributes and overall satisfaction, there is a need of a flexible statistical technique that can model this kind of complex, nonlinear relationships. Bayesian networks seem to be an interesting alternative to quadratic regression models in this respect. However, to the best of our knowledge, Bayesian networks have not been applied nor evaluated in this area; hence our motivation for examination of Bayesian networks in the context of CS research.

1.2.4. Current problems and challenges in CS&L research

Apart from the above-mentioned requirements, we have been able to identify also other reasons for which we speculate that the application of Bayesian networks could turn out worthwhile in solving marketing problems. We will address these issues in this section.

At the current stage of research, there seems to exist an agreement among marketing scientists towards the fundamental processes that explain the CS&L phenomenon [Garbarino and Johnson, 1997]. However, despite the agreement on the fundamental topics, many issues still remain to be discussed and elaborated upon. Firstly, primary focus of research is the successful conceptualisation of the constructs. Various researchers use different definitions, and/or comprehend the constructs in a different manner; moreover, measurement instruments, even when the conceptualisation is the same, are remarkably different, let alone standardized. Secondly, due to the complexity of the phenomenon, the precise nature of interactions among various concepts still remains unrecognised, especially cause-effect relationships and their relative strengths. In this context, Bloemer and de Ruyter [1999] argue for instance: "... the direct relationship between customer satisfaction and loyalty has remained somewhat equivocal", or with respect to the causal ordering between service satisfaction and service quality [de Ruyter *et al.*, 1997]. Thirdly, deeper insight is required in the effects of situational and contextual factors that influence consumer behaviour.

Besides the overall imperative to continue the CS&L research, there exist some difficulties that often restrain the CS&L research. In our opinion, these difficulties can be attributed to: i) theory representation problems, ii) existing modelling techniques that are hardly predictive, iii) application needs, and, iv) other limitations posed by existing approaches to theoretical modelling. Next, these concerns are briefly described.

i) Theory representation problems

> *"Marketing models are invaluable for the accumulation of the generalizable marketing knowledge"* [Van Bruggen and Wierenga, 2000].
>
> *"Marketing effects are caused by multiple variables and the relationship between cause and effect tends to be probabilistic"* [Malhotra, 1993].
>
> *"Decision [i.e. Structural Equation Modelling] models' claimed 'insight into marketing effectiveness' mostly seems to assume that a regression equation implies causation"* [Ehrenberg *et al.*, 2000, p.150].

Currently, the most prevailing empirical attempts to identify and model the CS&L theory processes are based on statistical methods of regression and covariance analysis [Hulland *et al.*, 1996; Baumgartner and Homburg, 1996]. These techniques include Structural Equation Modelling, Partial Least Squares, and multiple linear regression (for the review of these techniques see Section 3.3). Roughly speaking, it follows that the relationships among the psychological constructs taking part in the decision processes can be explained by pseudoindeterministic, usually linear, equations with coefficients interpreted as strengths of causal influence [Ehrenberg *et al.*, 2000]. Accordingly, assuming this approach is valid, humans tend to process information by linearly organized structures either implicitly or explicitly, with some random error term included.[1]

In CS&L research, on one hand, it is often assumed that the linearity of the relationships provide sometimes good approximations to nonlinear equations, at least within certain ranges; on the other hand, these approximations can often conceal relationships that would be significant if modelled at a nonlinear level [Rigdon, 1998]. Furthermore, it has also many times been shown that these relationships are not so simple and involve much more complexity [Mittal *et al.*, 1998].

We tend to think that the nature of causal relationships among latent constructs is much richer than just linear and should be explained by more expressive formalisms. As a matter of fact, we conjecture that failure to accept these requirements might lead to bias and impaired quality of research. In this regard, we suppose that Bayesian networks could be viewed as an alternative to curve-linear regression [Ping, 1996, see also Rigdon *et al.*, 1998].

Furthermore, our underlying assumption is that the processes through which consumers arrive at satisfaction and loyalty judgments tend to have a probabilistic nature [Jacoby and Chestnut, 1978; Malhotra, 1993], and the relationships among the cause and effects can take the form of any potential probability distribution. For instance, for a given focal construct, it could happen that depending on the value, or level, of an antecedent construct, the relationship between this focal construct and the antecedent might get stronger or weaker, or have a different character. We conjecture that such an approach can contribute to more successful development of the theory of Customer Satisfaction & Loyalty.

The contribution of Bayesian networks in this respect could be that they offer both the possibility of accounting for a rich nature of relationships and enable modelling of uncertainty explicitly using expressive probability formalism and theoretically sound probability calculus. Yet, they have not been examined on this issue in the context of the CS&L phenomenon.

[1] To be precise, recently, new methods have been proposed within the SEM approach that allow for modelling non-linear effects [e.g., Ping, 1996].

ii) Existing modelling techniques are hardly predictive and explanatory

"In the end the results of marketing science can be used to support marketing decision making in companies" [Van Bruggen and Wierenga, 2000].

"Compared to other modelling techniques, Structural Equation Modelling is more focused on explaining marketing phenomena than on predicting specific outcome variables" [Steenkamp and Baumgartner, 2000].

Marketing models can contribute to marketing management decision-making in companies in two ways: 1) indirectly, over time via empirical generalization of marketing knowledge from patterns by means of descriptive models [e.g., Ehrenberg *et al.*, 2000], 2) directly, in case of predictive and normative models, as well as by use of marketing management support systems [Van Bruggen and Wierenga, 2000].

However, the central theme of criticism towards existing dedicated prescriptive or decision models, such as Structural Equation Models, concerns their virtual lack of successful predictive capabilities and can be, quoting Ehrenberg *et al.* [2000], attributed to: 1) making little or no use of the large amount of well-established descriptive knowledge that exists, 2) their complexity and requirement of many parameters, 3) having no solid track-record of predictive practical applications, and 4) making unsubstantial causal assumptions.

More importantly, we argue here that SEM modelling does not excel in both predicting and explanatory power. Steenkamp and Baumgartner [2000] argue that SEM modelling is more focused on explaining phenomena than on predicting specific outcome variables. The predicting power of SEM is indeed meagre: it is not possible for any case to determine the value of latent variable by any means [e.g., Rigdon, 1998]. This limits the use of SEM in practice drastically, as managers would be most likely interested in these latent variable scores in the first place [*idem*, p. 278].

Moreover, besides the virtual lack of predicting capabilities, SEM modelling, in our opinion, can hardly be accepted also as an explanatory technique. The majority of explanation comes from the model specification procedure of defining and ordering latent variables and asserting causal relationships [Blodgett and Anderson, 2000]; once the SEM model is estimated, its explanatory potential is questionable.

It follows that there exists a need of a modelling methodology capable both of describing phenomena by making use of the existing descriptive theoretical knowledge, and making reliable predictions and prescriptions at the same time. The Bayesian networks formalism should be considered as an alternative to SEM modelling in this respect, since, in general, it enables making predictions, as well as allows for explanation. It can be expected that models that fulfil those conditions can contribute to the success of development of marketing scientific knowledge and thus to the direct and indirect adoption of marketing science.

iii) Application needs

> *"We propose that future models provide managers with what-if simulation capabilities"* [Leeflang *et al.*, 2000].
>
> *"What-if analysis have been widely applied in marketing and might very well be useful in analysing the impact of customer satisfaction, value, and price decisions"* [Rust *et al.*, 2000, p.448].
>
> *"Especially, it would be desirable to make decision models directly usable by managers instead of depending on staff studies or consulting projects as they do now. This will require better input modelling, more efficient computational methods, and the development of appropriately easy-to-use interfaces"* [Little, 1994].

It is a known fact that the adoption of CS&L theory by companies in practice is rather limited [e.g., Roberts, 2000; Simon, 1994]. Managers tend to rely on their experience and intuition, and rarely base their decisions on empirical research findings accumulated in descriptive models [Leeflang *et al.*, 2000]. However, it is also known that often much better decisions can be achieved by the use of marketing models [*idem*]. Yet better results are expected from the use of marketing management decision systems [Wierenga *et al.*, 1994; Wierenga and Van Bruggen, 2000]. In particular, a successful marketing management decision support system should be capable of providing answers to the following analytical queries [*idem*]: 1) "what" questions, 2) "why" questions, 3) "what-if" questions, 4) "what-should" questions.

Another important point related to the actual failure of existing decision models offered today is that the desirable software implementation should satisfy a number of criteria, some of which are: 1) completeness, 2) simplicity, 3) adaptability, 4) robustness, and 5) evolutionary [Little, 1994], and we are not aware of any technique that successfully fulfils these criteria.

The challenge therefore is to propose a modelling technique enabling adoption of the CS&L theory by providing managers with a desirable software implementation of decision models equipped with the capabilities suggested above, and eventually with a management decision support system [Wierenga and Van Bruggen, 2000]. The delivery of such a methodology will increase impact on managerial practice [Roberts, 2000].

The potential contribution of Bayesian networks in relation to the application needs in CS&L research is that they offer modelling capabilities useful in the marketing practice, for instance "what-if" simulations, probabilistic reasoning, etc.

iv) Other limitations posed by existing approaches to theoretical modelling

The quality of the theoretical research on Customer Satisfaction and Loyalty is also influenced by other limitations of the traditional techniques applied in the CS&L research today. Moreover, the adoption of marketing science is also sometimes hindered by various deficiencies and inconveniences related with the use of these techniques.

These include, for instance, the requirement of multivariate normality of data in case of Structural Equation Modelling. This assumption is typically violated in CS&L studies and can lead to seriously biased results [Rigdon, 1998; Hulland *et al.*, 1996]. Furthermore, some techniques that would not require any specific assumptions with respect to data distribution, would indeed require a particular type of data, i.e., only categorical, interval, or numerical data. Yet other deficiency is that they cannot handle missing data. Last but not least, they cannot facilitate optimal use of all available data in one model in situations, when two or more traditional models would be necessary. We elaborate more on these and other limitations of existing techniques, including SEM models, PLS models, and regression, in Chapter 3.

To be precise, most of these issues are well recognized in the modelling literature and various measures have been proposed to ease them. However, these measures are typically very complicated and require expertise in matrix algebra or advanced statistics. For instance, nonlinear structural equations are still cumbersome to estimate [Steenkamp and Baumgartner, 2000]. Furthermore, some of these limitations result from the underlying methodological inherently present in these techniques, and as such cannot be tackled by any means.

All these problems limiting the use of techniques in practice, as well as others not listed here, are absent or can be in principle more easily tackled with the Bayesian network approach. The contribution of Bayesian networks can be in this respect not only that the researchers' work can be more objective and easier, but also it will likely stimulate more widespread use of theoretically sound models in business practice.

1.3. Objectives of thesis

The primary objective of research presented with this work is to provide a critical evaluation of the application of Bayesian networks in theoretical and practical CS&L research, and propose new methods and solutions within the Bayesian network modelling to improve its current abilities with respect to specific requirements in theoretical and practical CS&L research.

In order to achieve this overall goal, we have designed a research strategy consisting of four case studies, in which we apply Bayesian networks in different settings and for different specific purposes. We will discuss the design of the case studies more precisely in the next section, and now we will specify the overall objective in terms of a number of more tangible objectives, each of which is particularized in terms of even more specific sub-goals. All these goals can be seen as pertaining to the use of Bayesian networks either in theoretical CS&L research or in practical CS research.

<u>I. Objectives with respect to the use of Bayesian networks in theoretical CS&L research:</u>

Table 1.3.1 presents the comprehensive list of all objectives and sub-goals regarding the use of Bayesian networks in theoretical CS&L research. The checkmark in the columns at the right hand side indicate the number of the case study, in which the sub-objective is addressed.

	Case	
	1	2
1. How can marketing theories be discovered (developed) by means of the Bayesian networks approach?		
a. Examine Bayesian networks in different scenarios:		
i. inductivist	√	
ii. deductivist		√
b. Propose and evaluate new methods for handling structural and measurement models, in particular aiming at:		
i. accounting for the measurement model		√
ii. latent construct validation		√
iii. finding the best dimensionality of latent constructs		√
c. Examine and discuss specific issues in theory development:		
i. the ability for modelling of moderating effects	√	
ii. the issue of accounting for mediating variables	√	
2. To what extent are purported marketing theories discovered with Bayesian networks subject to scientific justification? How can they be scientifically justified (validated)?		
a. Evaluate descriptive, predictive and explanatory potential of Bayesian networks on the example of the e-satisfaction and loyalty domain	√	
3. What is the added value of modelling marketing problems with Bayesian networks?		
a. Demonstrate the ability of performing probabilistic reasoning (forward, backward, inter-causal) in the domain	√	
b. Show the potential of performing what-if simulations	√	
c. Illustrate the potential of combination of prior knowledge with data	√	√
4. What are the strengths and weaknesses of Bayesian networks in terms of specific technical and statistical modelling issues, such as data distributional assumptions, missing data handling, etc	√	√

Table 1.3.1 Objectives of the thesis in the part on the theoretical CS&L research and the case study, in which the objective is achieved.

The first research question belongs to the context of discovery in marketing research. In turn, the second research question belongs clearly to the logic of justification. These two questions are central for the recognition of Bayesian networks as a legitimate technique in theoretical marketing research. The third question that deals with theoretical research aims to find out what is the added value of Bayesian networks in this respect. Finally, we will investigate the strengths and weaknesses of the Bayesian network approach from the perspective of theoretical CS&L research. This latter research question will be tackled by making the relevant observations wherever possible throughout the whole thesis.

<u>II. Objectives with respect to the use of Bayesian networks in practical CS studies:</u>

With regard to the practical CS research, we have proposed one research question that generally aims at evaluation of Bayesian networks in this stream of marketing research. We will also use the discussion in this part to pinpoint the strengths and weaknesses of the Bayesian network approach from the perspective of practical research in the light of other techniques currently used.

Table 1.3.2 contains the list of objectives and sub-objectives concerning the se of Bayesian networks in the practical CS studies.

	Case	
	3	4
1. How can Bayesian networks be applied in service feature/dimension importance/performance study?		
a. Adapt and examine Bayesian networks in classification of service dimensions analysis, aiming at in particular		
i. identifying the derived importance of service dimensions for overall (dis)satisfaction judgments	√	
ii. supporting marketing decisions by means of importance/performance analysis	√	
iii. discovering interaction effects (synergy and negation) among service dimensions	√	
b. Adapt and examine Bayesian networks in classification of service features (attributes):		
i. to evaluate the mediated model of overall satisfaction based on the technique of parent divorcing in the analysis of feature importance		√
ii. to find out whether in the mediated model, it is possible to treat satisfaction with service dimension as a hidden node, and thus optimise a questionnaire by not asking about satisfaction with service dimension		√
iii. to evaluate the noisy-OR model of overall satisfaction in the analysis of feature importance		√
2. What are the strengths and weaknesses of Bayesian networks in terms of specific technical and statistical modelling issues, such as data distributional assumptions, missing data handling, etc	√	

Table 1.3.2 Objectives of the thesis in the part on the practical CS research and the case study, in which the objective is achieved.

Because the case design listed above is somewhat complicated, we will elaborate more upon it in the next section and explain how we organize the flow of discussion to meet the objectives of this work.

1.4. Research strategy

We found that the overall goal of this work can be best realized with a research strategy consisting of case studies, each of which has its own background, data and objective that altogether make it possible to realize each of particular sub-goals listed above.

Consistently with the two views on CS&L research taken in this work, namely theoretical and practical, we designed in total four cases studies, two for each of these two views (see Tables 1.3.1-2).

In the first part, we consider various essential issues in discovery and development of marketing theories and evaluate Bayesian networks in the context of each of these issues. Therefore, the first two studies could be regarded as typical examples of theoretical CS&L research. The background of the first case study, set in the customer e-loyalty research context, is best suitable to discuss in what sense a Bayesian network model can be identified with a scientific theory (cf. objective 2.a), and to evaluate Bayesian networks in the context of accounting for moderating and mediating effects, as well as to demonstrate the ability of reasoning and what-if simulations.

The background of the second case study is especially suitable for the purposes of introducing latent constructs into modelling with Bayesian networks, as it contains data, in which some constructs are operationalized by means of multiple indicators. This study is set up in the traditional customer satisfaction and loyalty research context.

The second part in this work deals specifically with practical Customer Satisfaction research. The main objective of the first case study in this part is to develop and present a procedure for importance/performance analysis by means of Bayesian networks, so the customer survey data that we use are appropriate.

The unique feature of the second study in this part is that we make use of a customer survey in which satisfaction at the feature level as well as at the dimension level is measured. Most importantly, the customer survey that we use in this case study makes it possible to investigate whether asking respondents for satisfaction with service dimension is necessary or whether it is not necessary. To solve this question, we evaluate whether satisfaction with service dimensions can be estimated in a Bayesian network model parameterised only on basis of observed data on satisfaction with features and overall satisfaction.

We shall outline the objectives and their operationalization for each case study here briefly.

1.4.1. Case study 1

The first case study is set entirely in the on-line setting and so e-customer loyalty provides the application background for all our considerations in this study.

The first overall research question that we address in this study is how marketing theories can be developed by means of Bayesian networks. In this study, the question is first of all operationalized by examining Bayesian networks in the inductive approach (see the objective 1.a.i in Table 1.3.1). The study is an attempt to shift from isolated, web site specific findings to more generalized overall theory of e-loyalty. This is achieved by first "learning" four specific models from data describing visitors of four different portals, and then by constructing an overall model of e-loyalty by the examination of the probabilities of various dependencies in these models in line with the probabilistic framework.

The second research question that we investigate is the extent to which purported marketing theories discovered with Bayesian networks are subject to scientific justification. In this context, we examine how purported marketing theories discovered with Bayesian networks can be scientifically justified, or, in other words, how do we know that they can be recognized as a legitimate marketing theory. For this purpose, we evaluate descriptive, predictive and explanatory potential of Bayesian networks on the example of the e-satisfaction and loyalty domain (cf. objective 2.a). By doing so, we investigate whether the application of the Bayesian network approach can contribute to the understanding of the e-loyalty phenomenon by its purported ability of description, prediction, and explanation. We will evaluate each of these three requirements of any theory with the criteria recommended by the modern empirical orientation in the philosophy of science [Hunt, 1991], namely, with regard to the explanation of the e-loyalty, we will try to find out for example why some web users are loyal, or why some users have favourable attitude towards the website, etc. To examine the explanatory power of the Bayesian network models in a more systematic way, we will show that the e-loyalty phenomenon to be explained was indeed, by means of these models, somehow expected to occur. We will also evaluate whether it is subject to intersubjective certifiability; we will assess its pragmatism, and, last but not least, we will address the issue of empirical contents and testability [Hunt, 1991]. To test the adequateness of the prediction, we will make use of the models as predictive systems, and assess their predictive accuracy in comparison with other standard methods of prediction. Finally, the examination of the descriptive potential will be discussed in terms of probabilistic independencies between variables implied by the model, and the marginal probabilities of variables. It is worth noting that the achievement of the objective can be seen as a significant contribution of this work into the Bayesian network modelling literature.

Important requirements of techniques aiming to contribute to the scientific understanding of marketing phenomena, and e-loyalty in particular, are the issues of moderating effects and mediating variables [Sekaran, 1992; Bagozzi, 1994]. We will discuss and evaluate the capabilities of Bayesian network modelling in this context too (cf. objectives 1.c.i and 1.c.ii).

Third research question to be considered in this study is "what is the added value of modelling marketing problems with Bayesian networks?" If we claim that the Bayesian network approach can fulfil the supply-demand gap of marketing modelling, we have to explore the added value of this approach for modelling marketing problems. We acknowledge that the added value could be best revealed in a competitive setting by comparing our approach with other techniques; however, since the main objective in the thesis is the internal validation of the Bayesian network approach, we do not need to make any empirical comparisons with other techniques. In this sense, in our opinion, the added value can manifest itself in the ability of performing probabilistic reasoning and belief updating in the domain, and more specifically the potential of forward, backward, and inter-causal reasoning. In this case study we will demonstrate this ability (cf. objective 3.a). Furthermore, we will show the potential of performing "what-if" simulations (cf. objective 3.b), which should be seen as the second element of the added value of Bayesian networks.

The next objective is to illustrate the potential of combination of prior knowledge with data at hand (cf. objective 3.c). This objective can be seen as the part of the research question presented as item 3 in the list, namely what is the added value of modelling marketing problems with Bayesian networks.

Finally, simultaneously with the flow of discussion, we will also, whenever appropriate, attempt to address the fifth research question, and investigate what the strengths and weaknesses of Bayesian networks are in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc (cf. objective 4).

1.4.2. Case study 2

The second case study is set in the traditional customer satisfaction and loyalty setting in the service industry. In this study we continue examining Bayesian networks in its potential of development of CS&L theory expressed in the first research question, but in this study we take the deductive-like approach (cf. objective 1.a.ii in Table 1.3.1). In our implementation of the deductive approach, we depart from a position in which we first propose a series of competing theoretical hypothetical models, each partly supported by the extant CS&L literature; then we test each of these Bayesian network models empirically against the response data, and we accept the model that best fits the data as the model representing the true theory. This approach is also known in the marketing literature as the multiple competing hypotheses approach [e.g., Armstrong *et al.*, 2001], and, in relation to the dominant hypothesis approach and the exploratory approach, has been advocated as an important measure of improvement of the marketing science [Brodie and Danaher, 2001].

In this study, we introduce the issue of structural (latent construct) models and the measurement models in the CS&L research with the Bayesian network methodology in a more principled manner. A question that arises in this situation is how to account for latent constructs in a Bayesian network model, and how to empirically test models with latent constructs. The second main objective in this study is to investigate and demonstrate how to link the structural model, i.e. the theoretical model of relations between latent constructs, with the measurement model, i.e. the way that latent constructs are measured and presented in the model. To this end, we propose and evaluate new methods for accounting for the measurement models in Bayesian network modelling by using local Naïve Bayes structures (cf. objective 1.b.i). We show how a hidden network model can be parameterised, and evaluated in terms of its posterior probability. Whether our approach can be deemed successful, we will judge on the basis of theoretical outcomes of the most likely model, like the nature and strengths of relationships between constructs in the structural model and by examining the relationships in the measurement models. Furthermore, we will compare our approach with the approach applied today, which is based on taking the arithmetic mean of the indicator variables and using this value as observed variable; this comparison will be based mainly on using the models as classification systems.

As the third objective, we have proposed and examined a method of construct validation within the Bayesian network technology (cf. objective 1.b.ii). The construct validation approach taken in this study aims to assess whether the indicator variables relate to one potential construct, or to more constructs.

Fourth objective of this study is demonstrating the use of Bayesian networks for finding the best dimensionality of latent constructs (cf. objective 1.b.iii). Here a dimensionality is understood as the most likely number of values that a latent construct takes on. The assumption that underlies this objective is therefore that a concept under consideration is not continuous with respect to its scale of values, but it is rather discrete with only several potential values. Again, we show how this objective can be realized within the Bayesian network approach.

At last, with regard to the investigation into the added value of modelling marketing problems with the Bayesian network approach, we illustrate furthermore the potential of combination of prior knowledge with data at hand (cf. objective 3.c).

Finally, we will investigate the strengths and weaknesses of Bayesian networks in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc. (cf. objective 4). This sub-goal will be achieved throughout the case study by making observations whenever appropriate.

1.4.3. Case study 3

In the third case study, which is set in the traditional phone services, we focus specifically on problems facing marketing managers in relation with the performance of their products/services. The study concerns thus the use of Bayesian networks in practical CS&L research (cf. Table 1.3.2).

In this study, we address the issue of determining service factors driving satisfaction and/or dissatisfaction [Oliver, 1996; Hill and Alexander, 2000], and develop a methodology that will help identify the impact of satisfaction with service dimensions on overall satisfaction with the service. This impact can be best described by means of a classification scheme that we develop in terms of driving overall satisfaction and/or dissatisfaction. More precisely, our procedure enables classifying service dimensions as exciters, basic, satisfier/dissatisfier, or as a non-relevant dimension. By a service dimension, we will regard an aspect of the service consisting of specific features used to compare services with each other; hence, we will deal with dimensions such as customer service, sales service, billing, communication, etc.

Firstly, we adapt and examine Bayesian networks for the purpose of identifying the derived importance of potential factors for overall (dis)satisfaction judgments (cf. objective 1.a.i in Table 1.3.2). Our objective will be to find out which service/products dimensions are potential sources of (dis)satisfaction. To this end, we apply a procedure based on sensitivity analysis in Bayesian networks.

Secondly, our Bayesian network approach is evaluated for the potential of supporting marketing decisions by means of importance-performance analysis (cf. objective 1.a.ii). The objective of this analysis is to indicate these service dimensions on which the company should focus their resources in the first place, and which dimensions are objects of possible overkill. Some of the categories that we define are: low priority, action needed, opportunities, strengths, take care, and possible overkill.

The third topic we discuss is if and to what extent Bayesian networks can be applied for discovering of interaction effects among service dimensions (cf. objective 1.a.iii). We will adapt and examine the approach in this regard.

Finally, we will explore what are the strengths and weaknesses of Bayesian networks in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc. (cf. objective 2). In this regard, we will for instance demonstrate how Bayesian networks outperform other alternative methods applied in practical customer satisfaction studies by allowing for optimal use of all available data in one model.

1.4.4. Case study 4

In the case study no. 3, we illustrate the use of the Bayesian network approach for gaining insight into the nature of the relation between satisfaction at the service dimension level and overall customer satisfaction. The service dimensions are however difficult to control and manage in practice, because they usually encompass a wide range of specific and diverse service/ product features. The practical applicability of results of such studies is therefore limited. The predominant purpose of practical satisfaction research should thus be to evaluate the importance and performance of service/product features, rather than service/product dimensions, with relation to overall customer satisfaction. This assessment of the importance and performance boils generally down to the classification of the nature of relation between each feature and the overall satisfaction score and can be defined in a way that we proposed in the third case study. Recall that we have then defined four kinds of features' nature: sastisfier/dissatifier, exciter, basic, and non-relevant.

In order to facilitate classification of service/product features, in this study, we adapt and examine Bayesian networks in classification of service features. To do so, we propose two methods for reducing the complexity of the model with service features as parents of overall satisfaction.

The first technique, referred to as *parent divorcing*, consists in simplifying the model by introducing additional variables as effects of service features and parents of overall satisfaction and by making additional assumptions. In this way, we obtain a new model of overall satisfaction, that we will call a *mediated* model. These new variables in our model can presumably reflect customer satisfaction with a relevant service/product dimension, and are often operationalized in customer surveys. As the first objective, we aim thus to evaluate the mediated model of overall satisfaction based on the technique of parent divorcing in the analysis of feature importance (cf. objective 1.b.i in Table 1.3.2).

This approach suffers however from other imperfections. The most important of them is that it would require to operationalise the satisfaction at the dimension level by the questionnaire. This results in the extension of the question list by another six-ten questions. There are important aspects that affect the quality of satisfaction research when many questions are involved. For instance, Douglas [1995] argues that when a questionnaire is too long respondents get tired of answering questions, the phenomenon known as response fatigue, and are not willing to participate further. Furthermore, they tend to give uninvolved answers that are not a true reflection of their actual respondent's standpoint, which is another threat to the quality of the research. Last but not least, asking each additional question on a questionnaire is usually an extra cost for the company that orders customer satisfaction study.

Hence, the second objective is to find out whether in the mediated model, it is possible to treat satisfaction with service dimensions as hidden nodes, and thus optimise a questionnaire by not asking about satisfaction with service dimension (cf. objective 1.b.ii). In this new model, all the necessary parameters relating to hidden nodes are estimated on the basis of the remaining variables and their dependencies implied by the model.

In order to judge the applicability of this model with service dimensions as hidden nodes, we will perform two types of validation. In qualitative validation, we will compare the classification of the features using the scale developed in the previous study. In addition, the predictive accuracy will serve as the second type of validation.

The third objective is to evaluate the noisy OR-gate model of overall satisfaction for the analysis of feature importance (cf. objective 1.b.iii). The noisy OR-gate dependency model is generally a simplification of the former one, but on the other hand it presents several advantages to be discussed in more detail further. To operationalise this objective, we will assess the ability of performing classification of service features by counting the number of meaningfully classified features.

1.5. Relevance

The thesis is relevant both for a wide audience of researchers busy with buildings models from data, especially in the marketing and artificial intelligence research communities.

First of all, the work should be of primary interest to marketing academics that are busy with developing theoretical underpinnings for the CS&L phenomenon, and which are yet unfamiliar with Bayesian networks. They should find it especially stimulating to see a new, promising data analysis technique that can solve some marketing problems. Most importantly, they will find it very valuable to see how Bayesian networks can contribute to better understanding and explanation of the CS&L phenomenon, by a discussion of its potentials in explaining theoretical relationships among constructs, including the issues of moderating and mediating effects. This discussion will be accompanied by the detailed treatment of the added value, as well as strengths and weaknesses of the presented technique. Moreover, they will find it very useful and instructive to see how they can improve their research in terms of proposing managerial implications, and how the managerial practice can profit from the application of Bayesian network approach.

The work is also of a great importance for those marketing modellers who are not novice to the Bayesian network approach, as they will be vitally interested in finding out how one can account also for the measurement model, and what are the practical value and implications of the presented method of accounting for the measurement model. Having read this manuscript, marketing researchers will be able to decide for themselves whether they can improve and advance their scientific work by the application of Bayesian networks. Consequently, the manuscript in the part on the theoretical CS&L research can also be relevant for other social science researchers involved in causal modelling.

Next, the thesis is also very valuable for marketing practitioners who are concerned with low customer retention rates, and who strive to deliver more satisfaction with their service/products to customers. They will find the second part, i.e., the part on practical Customer Satisfaction research, especially relevant for analysing data they collect in customer satisfaction programmes. In this respect, marketing practitioners will get to know how they can determine the derived importance and performance of service features and dimensions. Furthermore, what is equally important, the relevance of the first part, more oriented on theoretical CS&L research, for practitioners cannot be overestimated as they will find out whether and how they can use theoretically sound models of CS&L in their managerial practice for improved decision support. We present also a discussion of unique capabilities of Bayesian networks, such as backward and inter-causal probabilistic reasoning, and "what-if" simulations. We expect that this discussion should have significant influence on modelling practice.

Finally, the work will be significant for computer scientists, statisticians, or econometricians interested in expert systems and data mining, as they might be interested in some technical issues arising in real-world applications of Bayesian networks and possible solutions, especially involving measurement modelling and sensitivity analysis in Bayesian networks. Last but not least, they will be interested in finding out whether and how Bayesian networks can be successfully applied in a new application domain, that is, in the CS&L research.

1.6. Remark on data used

We would like to stress that we are using existing, secondary data sets throughout the case studies. They all come from four different market research organisations operating in Belgium and the Netherlands that each performed independent customer satisfaction and loyalty studies. We would like to make it clear that we did not have any influence on specific issues concerning the questionnaire preparation and administration. Specifically, we had no influence on selection of theoretical constructs and other variables studied in these studies or on their respective measurement instruments. Furthermore, we are not in a position to give any details on the sampling procedure, e.g., how the survey respondents were selected and what was the response rate. This could bring along a number of consequences for the quality of the research, especially if its main objective was to deliver solid theoretical insights into the CS&L phenomenon.

Nevertheless, since our objective is primarily methodological, that is, to evaluate the Bayesian network technique rather than to draw conclusions concerning the theory or practice of Customer Satisfaction and Loyalty, we believe that the data sets we are using are appropriate enough for reaching the objective of this thesis. Certainly, the variables that we include in the case studies are typical of CS&L research and represent the main stream of CS&L research. More importantly, we suppose that the fact, that we could not control for the survey design and data collection procedure could eventually be in favour of our research strategy. Furthermore, we expect that if the results of our examination turn out positive, e.g., in terms of theoretical relationships between variables, then this fact could also be interpreted in favour of the Bayesian network approach because it would indicate the robustness of the approach with respect to data of worse quality.

1.7. Organisation of the thesis

The aims of the thesis will be fulfilled across various chapters in this thesis. In Table 1.7.1, we give a schematic overview of the topics we address in this dissertation. Chapter 2 introduces a formal description and the most important characteristics of the Bayesian network methodology. Chapter 3 gives a short review of the current Customer Satisfaction & Loyalty literature, and sketches existing causal modelling approaches applied today in CS&L research.

Chapter no.	Topics	Goal	
Chapter 2	- Definition of Bayesian network - Construction, validation, and use of BNs		
Chapter 3	- Review of the CS&L literature - Review of modelling techniques in CS&L research		
Chapter 4 (Case study 1)	- Inductive research - Discovering moderating effects - Accounting for mediating variables - Explanation, prediction and description - Probabilistic inference - What-if simulations - Combination of prior knowledge with data - Strengths and weaknesses	1.a.i 1.c.i 1.c.ii 2.a 3.a 3.b 3.c 4	Theoretical CS&L research
Chapter 5 (Case study 2)	- Deductive research - Latent constructs and measurement model - Construct validation - Dimensionality of latent constructs - Combination of prior knowledge with data - Strengths and weaknesses	1.a.ii 1.b.i 1.b.ii 1.b.iii 3.c 4	
Chapter 6 (Case study 3)	- Importance of service dimensions - Importance/performance of service dimensions - Interaction effects between service dimensions - Strengths and weaknesses	1.a.i 1.a.ii 1.a.iii 2	Practical CS studies
Chapter 7 (Case study 4)	- Classification of features - Optimisation of the questionnaire - The noisy-OR model of Overall Satisfaction	1.b.i 1.b.ii 1.b.iii	
Chapter 8	- Conclusions and final remarks		

Table 1.7.1 Organisation of the dissertation.

Discussion and the results of the use of Bayesian networks as tools for theoretical modelling in CS&L research are addressed in Chapter 4 and 5. In Chapter 4 we present a Bayesian network approach to theoretical research. The problem of the measurement model and latent construct validation in the context of CS&L research are the focus in Chapter 5.

The practical application of Bayesian network analysis in the practical customer satisfaction research and their evaluation is the main goal in Chapters 6 and 7. In Chapter 6 we introduce the importance/performance analysis of service dimensions as the traditional objective of practical satisfaction studies, and show and evaluate the use of Bayesian networks on the example of a real-world case study. The main theme in Chapter 7 is to investigate the importance/performance analysis applied to service features, and whether in the mediated model, it is possible to treat satisfaction with service dimension as a hidden node, and thus optimise a questionnaire by not asking about satisfaction with service dimension.

Chapter 8 concludes this work with final conclusions and implications of this research and suggests potential avenues for future work.

2. Bayesian networks

In this chapter we will introduce the underlying principles and formal assumptions in the Bayesian network modelling. We will also review some fundamentals of building and deploying Bayesian networks. The Chapter is organized in the following way. First, we present a short historical outline. Next, in Section 2.2, we give an exposition on probability and graph theory – two cornerstones of Bayesian networks. The definition of a Bayesian network is given formally in Section 2.3. Construction from domain knowledge is the topic in Section 2.4. Methods of learning from data are addressed in Section 2.5. Issues in validation are addressed in Section 2.6. How to use Bayesian network models is discussed in Section 2.7. In Section 2.8 we discuss briefly the articles that we found in the business/marketing literature whose objectives is the application of Bayesian networks. We close with conclusions resulting from this overview in Section 2.9.

2.1. Historical background

The origins of Bayesian networks date back long before the 1980s. Initially, their graphical semantics have been studied for long in mathematics and statistics but they have become applied more and more successful only in the 1990s thanks to recent developments advanced by the artificial intelligence research community. They have been introduced in the artificial intelligence as an alternative to existing expert systems. Bayesian networks were originally intended to be constructed purely from domain knowledge, but soon it became clear that the Bayesian network models could be also successfully constructed from data.

A short exposition on expert systems would be helpful at this place. An expert system can be broadly defined as "a computer system (hardware and software) that simulates human experts in a given area of specialisation" [Castillo *et al.*, 1997]. Two fundamental elements of any expert system are a knowledge base and an inference engine. The knowledge base (KB) incorporates definitions of data structures for a given domain (abstract KB), and actual data reflecting a concrete case at hand (concrete KB). The main purpose of the inference engine is to draw conclusions by applying the abstract knowledge to the concrete knowledge.

One of the most successful implementations of experts systems used to be rule-based systems. Rule-based systems use production rules of the form "if *condition* then *action*" as its abstract knowledge base, and forward chaining and backward chaining based on Boolean logic as its inference engine. Early rule-based systems were purely deterministic – they could deal only with problems that could be represented and solved in domains in which relations among entities were certain. Since the class of such problems is practically narrow, soon a need was recognised to augment the rule-base systems with some aspects of uncertainty.

In order to handle some aspects of uncertainty multiple approaches have been proposed. One of the most successful applications of rule-based expert systems augmented with certainty factors was MYCIN [Buchanan and Shortliffe, 1984]. Certainty factors were numbers from 0 to 1 attached to a rule and expressed in a sense a level of truth, or certainty, of the rule. As a result, rules in this system had the form "if *condition* then *action* with certainty *x*."

However, there were major problems involved with doing inference in systems with certainty factors. These problems were due to serious flaws in calculus with certainty factors, especially in tasks of combination and chaining [Jensen, 2001]. We first give an example of combining certainty factors. Imagine, we have two rules "if *a* then *c* with certainty *x*" and "if *b* then *c* with certainty *y*". Now, if we know both *a* and *b*, what should be the certainty of the fact *c*? The answer requires a function for combining certainties coming from two rules. Chaining can be explained in the following example. For instance, consider two rules: "if *a* then *b* with certainty *x*" and "if *b* then *c* with certainty *y*". Suppose we know *a*, then what is the certainty of *c*? Heckerman [1986] showed that any function for combination and chaining would, in some situations, lead to wrong conclusions.

In the search for mathematically and theoretically sound foundations for doing inference, the expert systems community has turned to statistics. The concept that seemed especially appealing was the Bayes' theorem, which became the cornerstone of the new generation of expert systems, because it enables combining new data with historical knowledge. Bayesian networks can be regarded as a kind of probabilistic expert systems. Other names frequently used are belief networks, Bayes' networks, Bayesian belief networks or causal networks.

Usually, one accepts the first half of the 1980s as the time in which Bayesian networks were introduced to the artificial intelligence community. It happened with the work of Judea Pearl entitled "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference" [Pearl, 1988]. The first real-world applications of Bayesian networks were MUNIN [Andreassen *et al.*, 1989] and Pathfinder [Heckerman *et al.*, 1992].

We have seen the bloomy days of Bayesian network in the 1990s, which, as we have mentioned, could take place thanks to development of effective algorithms for probabilistic inference and learning from data. The research on successful learning from data continues, and currently, it is one of the most active research areas within the machine learning community [Dietterich, 1997].

2.2. Probability and graph theory

2.2.1. Probability calculus

There exists more than one different calculus that deals with uncertainty in statistics. One of them is the classical probability calculus, and this is the one that is the foundation of further considerations in this dissertation. In this short overview of probability calculus we will make use of the set-theoretic definition of probability developed by A Kolmogorov in 1933 [Kolmogorov, 1933].

2.2.1.1. Basic axioms

Let us start with the basic notions in the probability calculus. It is convenient to perceive probability theory as dealing with experiments. Each experiments has a set of distinct *outcomes*. This set can be finite or infinite, but it must be well defined. The collection of all outcomes of an experiment is called the *sample space*. If the set of outcomes is finite we refer to the sample space as *finite*, and to every subset of such a sample space as an *event*. We do not consider the case of infinite sample space in this short introduction to probability calculus. A subset that contains exactly one element is called an *elementary event*. We can now define a probability function as follows.

> *Definition* 2.1. Suppose we have a sample space Ω containing n distinct elements. That is,
>
> $$\Omega = \{e_1, e_2, e_3, \ldots, e_n\}.$$
>
> A function that assigns a real number $P(\mathrm{E})$ to each event $\mathrm{E} \subseteq \Omega$ is called a *probability function* on the set of Ω if it satisfies the following conditions:
>
> 1. $0 \leq P(\{e_i\}) \leq 1$, for $1 \leq i \leq n$.
> 2. $P(\{e_1\}) + P(\{e_2\}) + \ldots + P(\{e_i\}) = 1$.
> 3. For each event $\mathrm{E} = \{e_{i_1}, e_{i_2}, \ldots, e_{i_k}\}$ that is not an elementary event,
>
> $$P(\mathrm{E}) = P(\{e_{i_1}\}) + P(\{e_{i_2}\}) + \ldots + P(\{e_{i_k}\}) = 1.$$

The pair (Ω, P) is called a *probability space*. Now we can give the axioms of probability theory.

> *Theorem* 2.1. Let (Ω, P) be a probability space. Then
>
> 1. $P(\Omega) = 1$.
> 2. $0 \leq P(\mathrm{E}) \leq 1$ for every $\mathrm{E} \subseteq \Omega$.
> 3. For E and $\mathrm{F} \subseteq \Omega$ such that $\mathrm{E} \cap \mathrm{F} = \varnothing$,
>
> $$P(\mathrm{E} \cup \mathrm{F}) = P(\mathrm{E}) + P(\mathrm{F}).$$

The conditions in this theorem are known as the axioms of the probability theory, and have been first proposed by A. Kolmogorov in 1933.

2.2.1.2. Conditional probabilities

The basic concept in probability calculus, and thus in Bayesian network modelling, is the notion of conditional probability.

Definition 2.2. Let E and F be events such that $P(\mathrm{F}) \neq 0$. Then the *conditional probability* of E given F, denoted $P(\mathrm{E} \mid \mathrm{F})$ is given by

$$P(\mathrm{E} \mid \mathrm{F}) = \frac{P(\mathrm{E} \cap \mathrm{F})}{P(\mathrm{F})}. \tag{2.1}$$

It is easy to explain the fundamental rule with probabilities understood as ratios. Sometimes, the following expression

$$P(\mathrm{E} \mid \mathrm{F})\, P(\mathrm{F}) = P(\mathrm{E} \cap \mathrm{F}), \tag{2.2}$$

is called the *fundamental rule* of probability calculus.

We will now define the concept of probabilistic independence. Two events E and F are *independent* in the probabilistic sense if one of the following holds:

1. $P(\mathrm{E} \mid \mathrm{F}) = P(\mathrm{E})$, and $P(\mathrm{E}) \neq 0$, $P(\mathrm{F}) \neq 0$.
2. $P(\mathrm{E}) = 0$ or $P(\mathrm{F}) = 0$.

Consequently, E and F are independent if and only if $P(\mathrm{E} \cap \mathrm{F}) = P(\mathrm{E})\, P(\mathrm{F})$.

Another crucial concept in Bayesian networks is conditional independence. The events E and F are *conditionally independent* given the variable G if $P(\mathrm{G}) \neq 0$ and one of the following holds:

1. $P(\mathrm{E} \mid \mathrm{F} \cap \mathrm{G}) = P(\mathrm{E} \mid \mathrm{G})$, and $P(\mathrm{E} \mid \mathrm{G}) \neq 0$ and $P(\mathrm{F} \mid \mathrm{G}) \neq 0$.
2. $P(\mathrm{E} \mid \mathrm{G}) = 0$ or $P(\mathrm{F} \mid \mathrm{G}) = 0$.

Now, another very useful law in probability calculus will be discussed. Suppose we have n events $\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_n$ such that $\mathrm{E}_i \cap \mathrm{E}_j \neq \varnothing$ for $i \neq j$ and $\mathrm{E}_1 \cup \mathrm{E}_2 \cup \ldots \cup \mathrm{E}_n = \Omega$. Such events are referred to as *mutually exclusive and exhaustive*. Finally, *the law of total probability* says that for any other event F, we have

$$P(\mathrm{F}) = \sum_{i=1}^{n} P(\mathrm{F} \cap \mathrm{E}_i). \tag{2.3}$$

Often, if $P(\mathrm{E}_i) \neq 0$, then the law is presented in the following way:

$$P(\mathrm{F}) = \sum_{i=1}^{n} P(\mathrm{F} \mid \mathrm{E}_i) P(\mathrm{E}_i). \tag{2.4}$$

2.2.1.3. Bayes' Theorem

Next, we will present the key theorem, which has been used for decades to compute the conditional probabilities of interest from known probabilities.

Theorem 2.2. Given two events E and F such that $P(\mathrm{E}) \neq 0$ and $P(\mathrm{F}) \neq 0$, we have

$$P(\mathrm{E} \mid \mathrm{F}) = \frac{P(\mathrm{F} \mid \mathrm{E}) P(\mathrm{E})}{P(\mathrm{F})}. \tag{2.5}$$

Furthermore, given n mutually exclusive and exhaustive events $\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_n$ such that $\mathrm{P}(\mathrm{E}_i) \neq 0$ for all i, we have for $1 \leq i \leq n$,

$$P(\mathrm{E}_i \mid \mathrm{F}) = \frac{P(\mathrm{F} \mid \mathrm{E}_i) P(\mathrm{E}_i)}{\sum_{j=1}^{n} P(\mathrm{F} \mid \mathrm{E}_j) P(\mathrm{E}_j)}, \tag{2.6}$$

This theorem is known as Bayes' Theorem, because it was originally developed by Thomas Bayes. It was published in 1763. We will also likewise refer to both of the equations.

Sometimes $P(\mathrm{F} \mid \mathrm{E})$ in formula 2.5 is called the *likelihood* of E given F and is denoted $L(\mathrm{E} \mid \mathrm{F})$. It is also informative to represent the Bayes' theorem as following:

$$P(\mathrm{H} \mid e) = \frac{P(e \mid \mathrm{H}) P(\mathrm{H})}{P(e)}, \tag{2.7}$$

which states that the belief we attach to hypothesis H upon obtaining evidence e can be computed by multiplying our previous belief in the hypothesis $P(\mathrm{H})$ by the likelihood $P(e \mid \mathrm{H})$ that e will be true if H is true. The probability $P(\mathrm{H} \mid e)$ is called posterior probability (or simply posterior), and $P(\mathrm{H})$ is called prior probability. The denominator $P(e)$ is often omitted in the considerations since it is merely a normalisation constant.

2.2.1.4. Random variables

One of the last concepts that we need to define is that of a random variable.

> *Definition* 2.3. Given a probability space (Ω, P) a *random variable* X is a function on Ω.

The *space* of X is the set of values that random variable can take. A random variable is called *discrete* if its space is countable or finite. For a random variable X we use $X = x$ to denote the set of all elements $e \subseteq \Omega$ that X maps to the value of x. In other words, $X = x$ represents the event such that $X(e) = x$.

We call $P(X{=}x)$ a *probability distribution* of the random variable X. If the variable is obvious from the context we an also write $P(x)$ instead of $P(X{=}x)$.

If the space of random variable X is a subset of real numbers then the *expected value* is given by

$$E(X) = \sum_x x P(x), \tag{2.8}$$

where $\sum_x$ means the sum as x goes over all values in the space of X.

Suppose we have two random variables X and Y defined each on the same sample space Ω. Clearly, they form a probability function on the Cartesian product of their spaces. Instead of referring to it as a probability function we call $P(X{=}x, Y{=}y)$ the *joint probability distribution* of X and Y, or simply *joint distribution*.

Finally, given a joint probability distribution $P(X{=}x, Y{=}y)$ we call the distribution $P(X{=}x)$ obtained by the summation as in the following expression:

$$P(X=x)=\sum_{y} P(X=x,Y=y) \tag{2.9}$$

the *marginal probability distribution* of X. We can say that variable X has been *marginalized out* of the joint probability distribution $P(X{=}x,\ Y{=}y)$.

2.2.2. Philosophical foundations of probability

From the philosophical point of view there exist two perspectives on probability that can be referred to as relative frequency approach, and Bayesian approach. The issue which one of these two approaches is superior in research in many disciplines is controversial and remains a unsolved question. We will touch on it very briefly.

The frequentist approach deals with the notion of probability as a relative frequency. A classical example of a probability as relative frequency is an experiment with tossing a coin. Each time the coin is tossed, the conditions are the same. More precisely, our knowledge is the same, but of course, there are some conditions we are not aware of as, the torque we out the coin on, the height, etc. In such repeated experiments, the relative frequency of each outcome (heads or tails) approaches a limit and this limit is called the probability of the outcome. Therefore, such a probability is called a relative frequency. It was formalized in 1928 by Richard von Mises [von Mises, 1928]. Proponents of this approach to probability are called frequentists.

The subjective, also called Bayesian, approach considers probability as a degree of belief. An example of this approach is when someone is asked to give her estimate of the chance of the total nuclear war. This probability is not a ratio, a relative frequency, or an estimate of a relative frequency. It is in fact a representation of one's subjective belief of the nuclear war given some actual political conditions in the world. This subjective probability approach is also called "Bayesian" because its proponents use Bayes' theorem to infer unknown probabilities from known ones.

The adjective "Bayesian" in Bayesian networks does not necessarily imply that the probabilities encoded by a Bayesian network model are subjective. As a matter of fact, they can be either; in this work however we will take the Bayesian approach to probability as dominating.

2.2.3. Graph theory

We will now very briefly address some of the most useful elements in the graph theory.

Suppose we have a set of possibly related objects $X = \{X_1, X_2, \ldots, X_n\}$. The set can be pictorially represented by a set of *nodes*, or *vertices*, each for one element in X. The nodes can be connected by lines, arcs or arrows, which are referred to as *links* or *edges*. If there is an edge between two nodes X_i and X_j we use L_{ij} to denote such a link. We will denote L as the set of all links.

Definition 2.4. A graph $G = (X, L)$ is defined by two sets X and L where X is a finite set of nodes $X = \{X_1, X_2, \ldots, X_n\}$ and L is a set of links, that is a subset of all possible ordered pairs of distinct nodes.

The links in a graph can be directed or undirected.

Definition 2.5. Let $G = (X, L)$ be a graph. When $L_{ij} \in L$ and $L_{ji} \notin L$, the link L_{ij} is called *directed*.

Now, we can define a notion of a directed graph.

Definition 2.6. A graph in which all the links are directed is called a *directed graph*. It follows that the order of the nodes defining a directed graph is important.

Subsequently, we need to define concepts of an adjacency set, a path between two nodes, followed by the definition of a closed path and a cycle.

Definition 2.7. Given a graph $G = (X, L)$ and a node X_1, the adjacency set of X_1 is the set of nodes directly attainable from X_1, that is, $Adj(X_1) = \{X_j \in X \mid L_{ij} \in L\}$.

Definition 2.8. A path from node X_i to node X_j is an ordered set of nodes $(X_{i_1} \ldots, X_{i_r})$, starting in $X_{i_1} = X_i$ and ending in $X_{i_r} = X_j$, such that there is a link from X_{i_k} to $X_{i_{k+1}}$, $k = 1, \ldots, r$-1, that is,

$$X_{i_{k+1}} \in Adj(X_{i_k}),\ k = 1, \ldots, r\text{-}1.$$

Definition 2.9. A path $(X_{i_1} \ldots, X_{i_r})$ is said to be closed if it has the same starting and ending nodes, that is, if $X_{i_1} = X_{i_r}$.

Definition 2.10. A cycle is a closed directed path in a directed graph.

Finally, one more definition we need is that of a directed acyclic graph. We can form it in the following way:

Definition 2.11. A directed graph is said to be *cyclic* if it contains at least one cycle. Otherwise, it is called a *directed acyclic graph*.

We will also refer to directed acyclic graph with an abbreviation "DAG", or "dag". There are other types of graphs possible, like partially directed graphs, etc. [Pearl, 1988], but we do not need make use of them in this dissertation, so their definition here is not necessary.

Finally, we will often refer to the nodes in a graph in terms of parents and children. Let us define these concepts too.

Definition 2.12. Given a directed graph $G = (X, L)$ and nodes X_i and X_j in X, X_i is called a *parent* of X_j, and X_j is called a child of X_i if there is a directed link from X_i to X_j.

2.3. Definition of a Bayesian network

Assume we have a set X=$\{X_1, X_2, \ldots, X_n\}$ of n random variables. We will now present a formal definition of a Bayesian network model.

Definition 2.13. A Bayesian network model, or simply a Bayesian network, is a pair (D, P), where D is a dag, and $P = \{p(x_1 \mid \pi_1), \ldots, p(x_n \mid \pi_n)\}$ is a set of conditional probability distributions, one for each variable, and Π_i is the set of parents of node X_i in D.

Thus, a Bayesian network consists of a network structure that encodes a set of conditional independence relations about variables in X and a set of local probability distributions associated with each variable. These two components define collectively the joint probability distribution for variables X in a domain. The nodes in D are in one-to-one correspondence with variables X.

There are several types of Bayesian networks depending on what kind of variables the model assumes. We will focus in this dissertation on the multinomial Bayesian network, in which every variable is discrete, i.e. it has a finite set of mutually exclusive states.[1] Another type are Gaussian Bayesian networks, which allow for continuous variables, that are parameterised by Normal distributions.

It follows from the definition that to each variable in a Bayesian network there is attached a table of conditional probability distributions. In multinomial networks in particular, the Cartesian product of possible values of parents form a number set of combinations. For each element in this set, that is, for each combination of parents' values, there is a separate conditional probability distribution. If a variable has no parents, then the table reduces to one unconditional probability distribution. Furthermore, the probabilities encoded by a Bayesian network may be physical or Bayesian.

One of the most appealing features of a Bayesian network model is that it allows for efficient representation of the joint probability distribution of all the variables that a model involves. This efficiency comes from the following theorem:

Theorem 2.3. Let BN be a Bayesian network over path $X = (X_1 \ldots, X_n)$. Then, the joint probability distribution $P(X)$ is the product of all local conditional distributions specified in BN:

[1] We would like to note that multinomial Bayesian networks can also model non-linear effects.

$$p(X) = \prod_{i=1}^{n} p(x_i \mid \pi_i) , \tag{2.10}$$

where π_i is the parent set of X_i.

The expression 2.10 defines the so-called *chain rule* in Bayesian networks [Pearl, 1988].

The factorisation of the joint probability distribution by the chain rule in Bayesian networks is the consequence of the fact that they admit Markov assumption. This means that each node is independent from its non-descendents given the value of its immediate parents. The set of nodes upon which all the remaining nodes are independent is called the Markov blanket of this node.

Our definition of Bayesian networks does not require that a model represents causal relations between variables in a domain. What is important is only that the relations of independence implied by a model admit the *d-separation* criterion in a domain [Pearl and Verma, 1990; Geiger *et al.*, 1990]. The d-separation criterion is a property of graphs that can be used to find out about the conditional independence relations between random variables. The idea of d-separation is based on the discrimination of three types of connections between children and parents in a network. These three types of connections are called *serial, diverging* and *converging* connections.

In Figure 2.3.1, we present the situation of serial connection between variables A, B and C. In this situation B is a parent of C, and A is a parent of B. If we have any evidence on A, then it will influence our belief about the variable B. New belief about B will in turn change our certainty about C. Similarly, if we know something about C, then it will cause us to change the certainty of B, and in consequence also A. However, if we know the value of B, then the fact that we know something new about A will not influence our certainty on C, nor vice versa. It follows that variables A and C are independent given B. We can also say that the connection between A and C is blocked. Otherwise, i.e., if B is not instantiated, the channel is active and any evidence on either A has impact on C and vice versa. As a result, we say that A and C are d-separated given B.

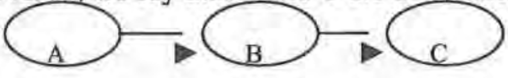


Figure 2.3.1 An example of serial connection.

In case of a diverging connection, which is exemplified in Figure 2.3.2, influence can be passed between any child variables unless the state of A is known. If the state of A is known we say that variables B and C are d-separated, and consistently, independent from one another.

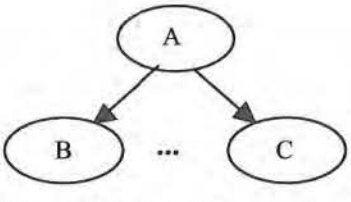


Figure 2.3.2 Diverging connection.

The third possible type of connection between three nodes is the converging connection. An example of this connection is presented in Figure 2.3.3. In this situation, the parent variables remain d-separated, and any specific knowledge with regard to the state of any one of them does not influence our belief about the distribution of the other one. However, if something is known about A, except for what is known from the knowledge of its parents, then information on one parent can have influence on other parents. In summary, in the converging connection, the causes are d-separated, unless something, in the form of soft or hard evidence, is known about A.

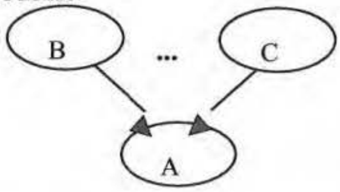


Figure 2.3.3 Converging connection.

In consequence, we have the following theorem:

> *Theorem* 2.4. Two distinct variables B and C are d-separated if for all paths between B and C, there is an intermediate variable A such that either the connection is serial or diverging, and A is instantiated, or the connection is converging, and neither A nor any of A's descendants received evidence.

The relation of the d-separation has the consequence that if A and B are d-separated with evidence e entered, then A and B are conditionally independent, i.e., $P(A \mid B, e) = P(A \mid e)$.

We must note that for faithful outcome probabilities it is not necessary that a model encodes true causal relations in a domain. Indeed, it is sufficient that a given network structure is equivalent with a true network in terms of the independencies it encodes. More precisely, we have the following definition:

> *Definition* 2.14. Let D_1 and D_2 be acyclic directed graphs over the same variables. D_1 is an I-*submap* of D_2 if all d-separation properties of D_1 hold also for D_2. If also D_2 is an I-submap of D_1, then D_1 and D_2 are said to be *I-equivalent* (independence-equivalent).

It is quite easy to check whether two Bayesian networks are I-equivalent. For this purpose it is convenient to define first a notion of a v-structure.

> *Definition* 2.15. An ordered triplet of nodes (X, Y, Z) is said to be a *v-structure* if and only if Z has converging arrows from X and Y, and there is no link between X and Y.

The following theorem is very important and helpful in determining if two different network structures encode the same set of independencies.

> *Theorem* 2.5. Two Bayesian networks are independence equivalent if and only if: a) they have the same associated undirected graph, and b) they have the same v-structures [Verma and Pearl, 1991].

2.4. Construction of Bayesian network models from prior knowledge

The construction of Bayesian network models can be based either entirely on the domain knowledge of the modeller, automatically resolved from a dataset, or can be a combination thereof. In case we resolve a Bayesian network from data, we refer to such a procedure as *learning* from data. In this section, we discuss how Bayesian networks can be constructed from prior knowledge, and in the section 2.5 we address the issue of learning from data. Once a BN model has been built, it can be queried for any (marginal) probability of interest.

Spiegelhalter *et al.* [1993] view the construction of a Bayesian network model as consisting of the three following components: qualitative modelling, probabilistic modelling, and quantitative modelling. In marketing modelling tradition however, building any econometric model whatsoever follows three steps: specification, parameterisation, and validation [e.g., Leeflang *et al.*, 2000]. Since in this dissertation we evaluate Bayesian networks as modelling tools in customer satisfaction research we will take this view and in what follows we will discuss how Bayesian network models can be built in marketing tradition.

2.4.1. Specification

The first step in building a Bayesian network model, like in any other model in general, should be a correct definition of the problem and identification of the goals of the model under consideration. In customer satisfaction research the dominant objective in construction of Bayesian network would be typically validation of a single hypothesis, a theory or some presumed knowledge. It could also be prediction or simulation. In Chapter 5 we present a case study to evaluate Bayesian networks in a model construction scenario in which we test some presumed hypothetical causal model in the CS&L research.

Once the objectives of modelling under consideration have been well defined, in the next stage the task is to identify and enumerate potential concepts, or variables, that may be relevant to the problem at hand [Heckerman, 1999].

What follows is the selection of concepts that are most worthwhile ones. This selection should be based again on domain knowledge.

In the next step, the modeller has to decide on each concept whether it can be best expressed with a continuous or a discrete variable. This distinction is important for the later use in terms of validation and performing inference. Since the focus in this dissertation is on multinomial Bayesian networks we assume that the variables are discrete.

Subsequently, defining the discrete nodes in terms of the fixed number of states running over an exhaustive and mutually exclusive state space should be the next task. The nodes with a fixed number of states can be either dichotomous (e.g., present/absent, true/false) or categorical (e.g., low/moderate/high). Furthermore, nominal as well as ordinal categorical variables can be incorporated.

In the next phase, we construct a directed acyclic graph that encodes assertions of conditional independencies among the variables included in the model. Defining the structure can be regarded as a specification or a parameterisation step – in the latter case arrows between nodes can be viewed as parameters.

The construction of directed acyclic graph could be guided by the assumptions of time ordering or causal knowledge [Heckerman, 1999]. The structure should be best captured in the way consistent with the joint probability space as implied by the chain rule of Bayesian networks (see formula 2.10).

During the construction of the structural dimension of the model, one can, if needed, adjust the definition of variables and consider introducing new variables. However, under some circumstances, e.g., when the number of conditional distributions is too large to be specified relative to the amount of data cases, the efficient construction of the model using the variables at hand is not feasible. As a result, the modeller is then advised to make use of special modelling techniques that simplify the structure. At some other times, the correct encoding of dependencies and independencies for the variables in question cannot be simply ensured. The techniques include, among others, the use of noisy-or and noisy-and gates, undirected relations, as well as divorcing [Jensen, 2001]. We evaluate some of these techniques in Chapter 7.

It must be noted that the dag provided by the modeller need not reflect true cause-effect relations in the domain in focus - the structure remains valid, which means that the dependencies and independencies reflect the domain, as long as the d-separation properties encoded by the dag hold for the domain, as there can be more than one different structures that represent exactly the same set of (in)dependencies. On the other hand, representation of causal knowledge is important, because it allows one to derive cause-effect statements about a domain after intervention, or manipulation. It is only in this latter case when the given network structure can be called a *causal* Bayesian network.

2.4.2. Parameterisation

Once the structure is provided, the next step in the construction of a Bayesian network model is the quantitative parameterisation. This task consists in specification of the numerical characteristics of the local dependencies by means of conditional probabilities. The probabilities are stored in conditional probability tables, usually called CPT's, in which the entries correspond to each state of a child node and all possible combinations of states for parent nodes.

Let us consider a small example of a Bayesian network shown in Figure 2.4.1. This network encodes a set of independencies between for random variables. The random variables A, B, C, and D in this example are all binary, and take values in the set $\{a_0, a_1\}$, $\{b_0, b_1\}$, $\{c_0, c_1\}$, and $\{d_0, d_1\}$, respectively. The conditional probability tables are shown next to the corresponding variables. Please note, that some entries in the CPT's are not shown, since it is obvious that the probabilities must sum up to unity.

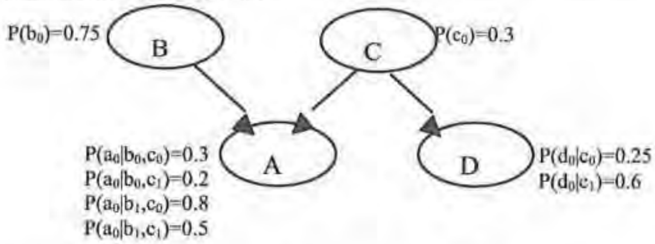


Figure 2.4.1 Structure and CPT's of an example of a Bayesian network.

The parameterisation of the Bayesian network in Figure 2.4.1 requires specification of 8 non-redundant parameters (probabilities).

In general, the parameters in a Bayesian network may be Bayesian or physical. The Bayesian (subjective) probabilities can come from prior knowledge acquired from experts, discussion panels, etc. If we have a database with cases we can estimate these probabilities from data. In the latter case the probabilities are physical. The issue of learning probabilities from data is discussed in more detail in section on learning from data.

Elicitation of subjective probabilities from experts can be accomplished by means of many techniques, of which a technique called *reference lottery* is quite popular or probability scale [see for instance Van der Gaag *et al.*, 2002]. Another technique is the *probability wheel* [Heckerman, 1999].

The next in the construction of Bayesian network models should be the procedure of validation. We discuss validation in Section 2.6, and now we will discuss construction of Bayesian networks from data sets.

2.5. Learning Bayesian networks from data

Learning Bayesian networks from data refers to automatic discovery of Bayesian network structures by means of computer programs. The two most popular approaches are the Bayesian approach and the constraint-based approach. They are the topics of our discussion now.

2.5.1. Bayesian approach

In case we have a database of cases at our disposal the estimation procedures vary as a result of existence of missing data in the database. The data in the database may be missing due to several reasons.

In line with the definition of Bayesian network, the task of learning from data, as typically presented in the BN literature, consists in determining the structure and estimating the local conditional probabilities. In practice, it is convenient to distinguish between four situations. In the first two scenarios, we actually know the structure and we need to estimate the probabilities: in one case when the data are fully observed, and in the other one when some data are missing. In the remaining two scenarios, we do not know the structure, so we need to learn both the structure and probabilities. In this case the method of choice also depends on whether the data are missing or not.

2.5.1.1. Known structure, full observability

We will start with a situation in which the structure is assumed known, so we will focus on learning the parameters (probabilities) from data. We begin with the situation when there are no missing data, and later proceed to the case with missing data.

In general, learning the parameters boils down to finding such values of these parameters that maximize some goal function. One of the most commonly used functions in this respect is the likelihood of the data. The likelihood of data is a function of data, or more specifically, a function of sufficient statistics, whereas actual conditional probabilities in the network and the network structure itself are assumed known.

By the chain rule, the likelihood of a single case $\mathrm{x}_l = \{x_1, x_2, \ldots, x_n\}$ is

$$p(\mathrm{x}_l \mid \Theta, B_s) = \prod_{i=1}^{n} \theta_{ijm} ,$$

where θ_{ijm} parameterises the probability that the variable i is in state m given the parent configuration π_{ij}, $\Theta=\{\theta_{ijm}\}$ is the set of all conditional probabilities associated with the network structure B_s.

If we let N_{ijk} be the number of cases in database $D=\{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N\}$ in which $x_i=k$, and $p_i=j$, then the likelihood of the entire data set D, assuming mutual independence of data and parameter independence is

$$p(D \mid \Theta, B_s) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{N_{ijk}}$$

where n is the number of variables, q_i is the number of possible parents' values combinations for the variable i, r_i is the number of states the variable i can take on.

The Maximum Likelihood (ML) configuration of parameters can be reached when we do not use any prior estimates of the parameters and therefore they are calculated as

$$\theta_{ijk} = \frac{N_{ijk}}{\sum_{k=1}^{r_i} N_{ijk}}, \tag{2.11}$$

where N_{ijk} is the number of cases in D, in which $X_i = x_{ijk}$, and $\mathbf{Pa}_{ijk} = \mathrm{pa}_{ijk}$.
The Maximum A Posteriori (MAP) configuration can be computed as

$$\theta_{ijk} = \frac{\alpha_{ijk} + N_{ijk}}{\sum_{k=1}^{r_i} (\alpha_{ijk} + N_{ijk})}, \tag{2.12}$$

The Maximum Likelihood estimate of the probabilities θ_{ijk} can be therefore seen as a special case of the MAP estimation, in which we are *a priori* completely uninformed and rely only on the data and let the priors be $\alpha_{ijk} = 0$.

2.5.1.2. Known structure, missing data

In principle, the problem of missing data can be solved by plugging all possible values in the place of missing data, where each plugged value has some probability of occurrence, and then make use of the total law of probability in Equation 2.4 and the scoring function 2.14. However, when the amount of missing data is large, this procedure is not feasible in practice.

When the structure is known and some data are missing, then the idea is typically first to find the best estimation of probabilistic parameters, and next to plug those values in some Bayesian measure of goodness of fit to score for the structure.

The case with missing values is more complicated than since the probabilities cannot be considered in apart from one another, since the estimates depend on each other when some data are missing.

The most frequently applied methods of parametric estimation in case of missing data include: i) Monte Carlo estimation [Heckerman, 1995], ii) EM estimation [Lauritzen, 1995], and iii) gradient learning [Russel *et al.*, 1995]. Of these methods, the most popular is perhaps the EM algorithm. Monte Carlo techniques are very precise, however they require large number of calculations, to converge to the local optimum. We address the EM optimisation thoroughly in Chapter 5.

2.5.1.3. Unknown structure, full observability

In practice, evaluating of every possible network structure is not feasible since the complexity of such a procedure grows exponentially [e.g., Cooper and Herskovits, 1992]. The number of possible belief network structures containing *n* nodes can be computed using the following recursive formula [Robinson, 1977]:

$$f(n)=\sum_{i=1}^{n}(-1)^{i+1}\binom{n}{1}2^{i(n-1)}f(n-1).$$

For instance, for n=2, the number of possible structures is 3; for n=5, it is 29.000, and for n=10, it is approximately $4.2\text{x}10^{18}$. As a result of the complexity, a method other than exhaustive enumeration of possible network structures is necessary. The last decade witnessed an active research in this field.

Consequently, there is a need for procedures that select only some candidate network structures and score them subsequently. Learning Bayesian networks can be thus seen as consisting of two elements: a search algorithm, and a score measure. As the score measure, we can use the Bayesian metrics discussed in Section 2.6.1. Here we will discuss the search algorithms. Usually, these search algorithms make successive changes in the structure by arc operations. Moreover, they employ the property of decomposability of the Bayesian metrics. The potential operations are arc removal, arc reversal, or arc addition. Of course, all changes are subject to the constraint that the resulting network remains acyclic.

Among those procedures the K2 algorithm [Cooper and Herskovits, 1992], simulated annealing, arc reversal, and algorithms based on genetic evolutionary are the most popular [Larranaga *et al.*, 1996]. Another method that avoids local maxima is hill-climbing [Chickering *et al.*, 1995].

We will give a short exposition of the greedy search, as this is the method that we will use later in this work. The greedy-search algorithm, known as K2 [Cooper and Herskovits, 1992], requires a prior ordering of variables as input. Nodes that come sooner on the ordering are tested as potential parents of the nodes that come later. The algorithm starts by making the assumption that a node has no parent. Alternatively we can start with a random graph, or a prior network. Following, it tests every node that appears prior in the ordering as a potential parent, and adds that node whose addition most increases the probability of the resulting structure. A potential problem of the greedy nature of the K2 algorithm is that it may be stuck in a local optimum.

2.5.1.4. Unknown structure, partial observability

The scenario in which the structure is unknown and data are missing is most difficult. In this case we need to direct the search and score it with a Bayesian metric. For a given model we can use again for instance the Cheeseman-Stutz approximation [Cheeseman and Stutz, 1995; Chickering and Heckerman, 1997] for models with missing data given as:

$$\log p(D\,|\,B_s)\approx\log p(D'\,|\,B_s)+\log p(D\,|\,\tilde{\Phi}_{B_s},B_s)-\log p(D'\,|\,\tilde{\Phi}_{B_s},B_s)$$

where D is the complete data, $\tilde{\Phi}_{Bs}$ is parameterisation of the model based on the ML or MAP estimation given data D, D' is completion of the data D given $\tilde{\Phi}_{Bs}$, B_s is the structure.

If we, in addition, allow for models with hidden variables then the Cheeseman-Stutz approximation for hidden variable models would be [Chickering and Heckerman, 1997]:

$$\log p(D \mid B_s) \approx \log p(D' \mid B_s) - \log p(D' \mid \tilde{\Phi}_{Bs}, B_s) + \frac{d'}{2}\log N +$$
$$+ \log p(D \mid \tilde{\Phi}_{Bs}, B_s) - \frac{d}{2}\log N$$

where d is the structural dimension, and d' is the effective dimension. The difference between these dimensions is a more subtle issue and requires more attention, therefore, we treat this topic in more detail in Chapter 5.

The BIC (Bayesian Information Criterion) approximation for hidden variable models is given as [*idem*]:

$$\log p(D \mid B_s) \approx \log p(D \mid \tilde{\Phi}, B_s) - \frac{d'}{2}\log N,$$

where $p(D \mid \tilde{\Phi}, B_s)$ is the likelihood of the model in the ML configuration of the model's parameters $\tilde{\Phi}$, N is the number of observations, and d' is the dimension of the model [Chickering and Heckerman, 1997].

2.5.2. Constraint-based approach

This approach is the basis for constraint-based learning of Bayesian networks in algorithms such as IC [Pearl and Verma, 1990], SGS and PC [Spirtes and Glymour, 1991], and FCI [Spirtes, Glymour, Scheines, 2001].

Probably, the most widely known instance of the constraint-based approach is the PC algorithm, implemented in a number of Bayesian network software packages. The input to the PC algorithm is a dataset. The algorithm performs a series of statistical tests for independence between the random variables. It starts with the marginal independence, and further proceeds with tests of conditional independence.

1) Form the complete undirected graph C,
2) Thinning – removing adjacencies in C by identifying independence relations:
    a. Zero-order (unconditional),
    b. First-order (conditional on 1 variable),
    c. Second-order (conditional on 2 variables),
    d. And so on...,
3) Orientation - identifying unshielded colliders,
4) Completion - directing the remaining links so that the resulting graph is a DAG.

Let us review the subsequent phases of the PC algorithm in more detail.

In Phase 1, we form the complete undirected graph C on the variables V, so that all nodes are connected with each other. There are thus no directed links at all at this moment.

Phase 2 can be viewed as thinning of the complete undirected graph C, formed in Phase 1. This phase is shown using a pseudo code in Figure 2.5.1. Beginning with a pair of variables, call it x and y, adjacent in C, we start considering dependencies, in which the conditioning set S is empty (cardinality of subset S of *Adj*(C, x)\{y} is 0, so it means that we test for marginal independence). If the variables x and y are independent, that is I(x, y), then we remove the edge between x and y in C and remember this fact as the separation sets Sepset(x, y) and Sepset(y, x). The graph C is thus constantly updated. We repeat this check until all ordered pairs of adjacent variables x and y such that cardinality of adjacency set *Adj*(C, x)\{y} is greater than or equal 0 have been tested for marginal independence. Next, we perform a similar procedure considering not marginal independencies, but independencies conditional on subsets S of *Adj*(C, x)\{y} of cardinality equal 1, 2, ... and so on. For each adjacent pair, we remember the separation sets Sepset(x, y) and Sepset(y, x). We stop when for each ordered pair of adjacent variables nodes x, y, the set *Adj*(C, x)\{y} has the cardinality less than *n*.

```
1°. n = 0;
2°. repeat
        - repeat
            - select 1) x,y adjacent in C such that |Adj(C,x)\{y}|
              >= n, and 2) subset S of Adj(C,x)\{y} of cardinality n;
            - if I(x, y|S), then delete edge x - y from C and record
              S in Sepset(x,y) and Sepset(y,x);
        - until all ordered pairs of adjacent x,y: |Adj(C,x)\{y}|
          >= n and all subsets S of |Adj(C,x)\{y}|=n have been
          tested for independence;
        - n = n + 1;
    - until for each ordered pair of adjacent vertices x, y, we
      have |Adj(C,x)\{y}|< n.
```

Figure 2.5.1 Pseudo-code representation of Phase 3 - thinning.

Phase 3 is the orientation phase. In this phase, the search for potential unshielded colliders is performed. A potential unshielded collider can be graphically shown in Figure 2.5.2.

$$\boxed{x} \;\text{———}\; \boxed{y} \;\text{———}\; \boxed{z}$$

Figure 2.5.2 Y is a potential unshielded collider.

For each such triple x, y, z, we orient x – y – z as x $\rightarrow$ y $\leftarrow$?z, if and only if y is not in the separation set Sepset(x, z).

Phase 4 can be seen as the completion phase, because now all the undirected links must become directed. Of course, the completion of directions must obey the elementary assumptions in the Bayesian networks that there may be no cycles between any pair of nodes, and that all links are directed. In the PC algorithm, this is achieved in two steps:

1) If
    a. $x \to y$,
    b. y and z are adjacent,
    c. x and z are not adjacent,
    d. and there is no arrowhead at y,

   then orient y - z as $y \to z$.
2) Furthermore, if
    a. there is a directed path from x to y,
    b. and there is an edge between x and y,

   then orient x - y as $x \to y$.

The basic underlying feature of the PC algorithm is thus the idea of an unshielded collider, because it is the core of the orientation of undirected links. All the remaining links are directed as a result of this observation only. As we can see, the initial ordering of variables is not of the critical importance in the PC algorithm, because the output is in fact a family of independence equivalent BN structures.

2.6. Validation

One can differentiate between four main criteria for validation of marketing models. These criteria are measure reliability and validity, face validity, statistical validity, and use validity [Naert and Leeflang, 1978; Lilien *et al.*, 1992].

Clearly, marketing models should satisfy all of these criteria as much as possible, but the priority should be given depending on the chosen method of validation. It should again reflect the intended use of the model. For Bayesian networks in particular, if the aim were to use for classification purposes, then the appropriate method of validation would be some measure of predictive quality. If we intend to use the network in an explanatory study, or the goal is justification of a theory, then the method of choice should be some Bayesian metric or constraint-based validation. We will present some methods of statistical validity of the Bayesian network models focusing on those that we use further in this dissertation. Statistical validity concerns some measure of goodness of fit of the model and the reliability of the estimated coefficients of the model [Lilien *et al.*, 1992].

2.6.1. Bayesian metrics

The validation based on Bayesian metrics aims at evaluating the posterior probability of the model structure. In the full Bayesian approach we treat the structure of the model as a discrete random variable whose states correspond to the possible network structure hypotheses B_s^h, (the subscript 's' stands for structure, 'h' for hypothesis) and for which we assess the prior probabilities $p(B_s^h)$. Then, given a random sample D, we compute the posterior distribution $p(B_s|D)$ and the posterior distributions $p(\Theta_s| B_s^h, D)$, and use these distributions to compute expectations of interest. For instance, to predict the next case x_{N+1} after seeing the database D, we have

$$p(x_{N+1} \mid D) = \sum_{B_s^h} p(B_s^h \mid D) \int p(x_{N+1} \mid \Theta_s, B_s^h) p(\Theta_s \mid D, B_s^h) d\Theta_s \, ,$$

where $p(\Theta_s| B_s^h, D)$, is the likelihood of the data given structure B_s^h and is easy to compute. This full approach requires the integration over all possible assignments of probabilities Θ_s for each candidate network structure.

The posterior probability $p(B_s|D)$ of the model structure B_s can be from Bayes' theorem expressed as

$$p(B_s \mid D) = \frac{p(B_s) p(D \mid B_s)}{p(D)},$$

where $p(D)$, i.e. the probability of data D, is a normalization constant that does not depend on the model structure. Thus, the posterior probability depends on the prior probability of the network $p(B_s)$, and the marginal likelihood of the data $p(D| B_s)$. So, when the models are assumed equally probable *a priori*, the network structure can be seen as valid when its marginal likelihood is maximal among all potential model structures. This approach would require that we know the marginal likelihood for each possible structure, but it is in practice not feasible.

In statistics, there are two approaches when dealing with probabilistic validation of models: model selection and selective model averaging [e.g., Heckerman, 1995]. Bayesian model selection aims to select one "good" model from among all possible models and use it as a if it were the correct model. This approach is sufficient when the selected model is remarkably more probable than its alternatives. This is often the case when we have little variables for which there is a lot of data [Heckerman *et al.*, 1999]. But the question how we decide if it is good "enough" is difficult to answer in theory [Heckerman, 1995]. Clearly, the closer the posterior probability to one, and the bigger the difference between the best and the second best model, the more confidence we have in the learned model. So, the decision whether to accept or reject the model is often of the subjective nature. The latter approach is to select a manageable set of good models from among all possible models and Bayesian model averaging is recommended when the likelihood values of two models are close.

There are several issues that we have to keep in mind when discussing the validation of models learned "automatically" from data, as we do in Section 3.5.1. Search methods, e.g., the greedy algorithm, do not evaluate every possible and legible structure that could be constructed from the variables under consideration, so in general, it might be true that some distinct model structures are more probable. There are two reasons for that. Firstly, the input to the search algorithms is an ordering on variables that in fact determines the direction of dependencies. A lot of potentially highly probable structures are thus not tested during the search for the best model. Secondly, for a given ordering of variables, the algorithms do not evaluate every possible combination of parents for any node. Recall from Section 4.3.8 that the greedy algorithm, for instance, begins by making the assumption that a node has no parents, adds incrementally that parent that most contributes to the probability of the resulting structure. The "greediness" of the algorithm can theoretically lead to the selection of not optimal parent configuration, for example when neither single one of two nodes *X* and *Y*, which are in fact the most probable parents of a node *Z*, is more probable than an other single third node *W*. In this situation, the greedy algorithm will select falsely the node *W* as the most probable parent and will never evaluate the nodes *X* and *Y* as common parents at all for *Z*. In practice however findings from other experiments with the greedy search have shown that the algorithm does well and often reaches optimal decisions [Chickering *et al.*, 1995].

It is convenient to make a distinction between the cases of complete and missing data.

2.6.1.1. Complete data

One of the most popular scores used to compare different Bayesian network structures is the marginal likelihood of the model structure B_s given the data. The meaning of this score is the likelihood of the data taken over all possible probabilistic parameterisation distributions Θ_s of the model B_s:

$$p(D \mid B_s) = \int p(D, \Theta_{B_s} \mid B_s) d\Theta_{B_s} \, , \tag{2.13}$$

Given a set of assumptions, the integral has a closed form solution. These assumptions are: multinomial sample, parameter independence (global and local), parameter modularity, the distributions are Dirichlet, and complete data [Heckerman *et al.*, 1995].

Let us consider the assumption of parameter independence in more detail. Global parameter independence says that the parameters associated with each variable in a network structure are independent. The assumption of local parameter independence says that the parameters associated with each instance of parents' configuration for a variable are independent. These assumptions were proposed by Spiegelhalter and Lauritzen [1990]. We can write them formally, respectively, as:

a) $p(\Theta_{Bs} \mid B_s) = \prod_{i=1}^{n} p(\Theta_i \mid B_s)$,

b) $For\ i = 1, \dots, n : p(\Theta_i \mid B_s) = \prod_{j=1}^{q_i} p(\Theta_{ij} \mid B_s)$,

where n is the number of nodes in the model B_s, and q_i is the number of all possible combinations of states of parents of the node i.

Given these assumptions, the formula for calculating the marginal likelihood was first derived by Cooper and Herskovits [1992] and looks in the following way:

$$p(D \mid B_s) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \cdot \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}, \qquad (2.14)$$

where r_i is the number of states of the node i, N_{ijk} are the observed counts of variable i in state k given the jth configuration of the node's parents, $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$, a_{ijk} is the prior precision, or the counts given by an expert as a priori estimates, and $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$. This score is often called the CH, or BD (Bayesian Dirichlet) metric.

The priors a_{ijk} reflect domain expert knowledge about the domain. The specification for all possible child-parent combinations would be unfortunately a tedious task. Therefore, some simplifying suggestions have been proposed. For instance, Cooper and Herskovits [1992] suggest a simple uninformative assignment a_{ijk} =1, for which the BD metric is known as the K2 metric. Another uninformative assignment was proposed by Buntine [1991] and equals $a_{ijk} = a_{ij}/(r_i q_i)$.

A very nice property of the marginal likelihood score in Equation 2.14 is that it lends itself to decomposition. It is easy to notice that the formula can be split into a product of factors

$$g(x_i, \pi_i) = \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}, \qquad (2.15)$$

where $g(X_i, \Pi_i)$ is the local contribution of a node X_i and its parents Π_i to the marginal likelihood of the entire model.

2.6.1.2. Missing data

An important consideration when treating missing data is whether or not we can ignore the process by which the data occur as missing. There can be generally two situations in which the data occur as missing: missing at random and missing systematically (or not at random) [Little and Rubin, 1987].

In the former case, when the data are missing at random, the missing-data mechanism is ignorable in the sense that the statistical inference is not dependent on it.

However, when the probability of a missing value for a variable depends on the unobserved, true state of this variable, then the missing-data mechanism is said to be not ignorable, and the data are missing systematically. The random sample is not representative of the population in any respect in this situation.

We will assume that the data are missing at random, since this assumption is easier to deal with. Unfortunately, in case of missing observations, the marginal likelihood score defined in Formula 2.14 cannot be applied and can only be approximated. There have been several approximated measures of the marginal likelihood proposed in the literature. The best and most frequently used approximations are the Cheeseman-Stutz approximation, abbreviated as the CS score, [Cheeseman and Stutz, 1995] and the Bayesian information criterion, or the BIC score [Schwarz, 1978]. See [Chickering and Heckerman, 1997] for derivation and a discussion of these measures. Here we will only present mathematic formulae that can be applied to obtain them. Note, that we use the logarithm of the actual approximation, as it prevents from computational flaws (the true marginal likelihood is a very low positive number), and is easier to perceive.

The Cheeseman-Stutz approximation is expressed in the following way:

$$\log p_{CS}(D \mid B_s) \approx \log p(D' \mid B_s) + \log p(D \mid \tilde{\Theta}_{B_s}, B_s) - \log p(D' \mid \tilde{\Theta}_{B_s}, B_s), \quad (2.16)$$

where D is the complete data, $\tilde{\Phi}_{Bs}$ is parameterisation of the model based on ML estimation given data D, D' is completion of the data D given $\tilde{\Phi}_{Bs}$, B_s is the model.

The Bayesian information criterion is expressed as:

$$\log p_{BIC}(D \mid B_s) \approx \log p(D \mid \tilde{\Phi}_{B_s}, B_s) - \frac{d}{2} \log N, \quad (2.17)$$

where $p(D \mid \tilde{\Phi}_{B_s}, B_s)$ is the likelihood of the model, N is the number of observations, d is the effective dimension of the model.

With the Bayesian approach it is possible to evaluate models with hidden nodes, i.e., variables whose values are absent in every case in dataset. This situation needs more attention and is covered in more detail in Chapter 5.

2.6.2. Constraint-based approach

Another approach to validation is based on the verification whether a given network structure admits relationships of conditional independencies between variables. We have discussed this approach in Section 2.5.2 on the constraint-based learning of Bayesian networks from data by means of algorithms, such as PC, IC or SGS [Spirtes *et al.*, 2001]. The underlying idea of identifying the independence relations by performing some statistical tests can be used to validate any Bayesian network structure hypothesized by an investigator. One of these methods is the TETRAD method, which tests the fit of models for categorical data in which all variables are recorded in the data [Scheines *et al.*, 1994].

2.6.3. Minimum Description Length approach

The approach based on Minimum Description Length principle comes from the coding theory, where the idea is to code a string with as few bits as possible. In the case of Bayesian networks this principle includes: 1) the length required to store the structure of the network, 2) the length required to store the parameters associated with the network, and 3) the length of the description of the dataset compressed using the Bayesian network structure and parameters [Lam and Bacchus, 1993; Bouckaert, 1993].

Asymptotically, the MDL approach is equivalent with the Bayesian Information criterion [Schwarz, 1978] that we discuss in section on learning Bayesian networks from data.

2.6.4. Predictive accuracy

As we have already mentioned, Bayesian network models can be used in classification tasks to classify a new case whose class is unknown. As its outcome a Bayesian network classifier yields the probability distribution of the class variable conditioned on what is known on other variables. Usually, the state that receives the most probability is assigned as the class. Indeed, the posterior probability distribution of the class variable depends actually only on the values of variables in the Markov blanket of the class variable. However, if the variables in the Markov blanket are not observed or not known, then the distribution will be determined on what is known by means of inference in the network. So, an advantage of Bayesian networks as classifiers is that it allows for missing values in the case to be classified.

In this dissertation, we consider Bayesian networks that in the Bayesian network classification terminology are referred to as generalized Bayesian network classifiers. There exist also other Bayesian network tailored specifically at classification tasks, as Naïve Bayes, tree augmented Naive Bayes, Bayesian multinets, and other [Friedman *et al.*, 1997; Cheng and Greiner, 1999].

Some of the most widely used measures of predictive accuracy are general performance, Brier score, and average logarithmic log loss. We will discuss them now in more detail.

2.6.4.1. General performance

The general performance is the most popular measure of the predictive accuracy. Classification accuracy score, also referred to as *percentage correct*, is defined as the percentage of correctly classified cases on the test set.

The drawback of the general performance metric is that it does not take the advantage of the uncertainty of the outcome of classification into account.

2.6.4.2. Quadratic (Brier) score

Among the best-known metrics of predictive accuracy is the Brier score [Panovsky and Brier, 1968; see also McLachlan and Peel, 2000 for a similar principle in the context of mixture models]. The intuitive idea behind the Brier score is that in case the posterior probability of a specific category of overall satisfaction is remarkably higher than for the other categories and the prediction is correct, then the quality of such a forecast is better as if the distribution of categories was more resembling uniform distribution. For each case in the test data the model gives a prediction of posterior probability distribution p_{ij} over the states j=1, ..., k of the class variable. The Brier score of this prediction can be defined as:

$$B_i = \sum_{j=1}^{k} (p_{ij} - s_{ij})^2$$

where s_{ij} = 1 if the predicted state is the same as the true state, and s_{ij} = 0 otherwise. Now, if the network would yield a correct prediction with full certainty, then the Brier score would equal to 0. On the other hand, if the network assigns probability 1 to a false state, the score would be 2. Consequently, for each case the Brier score can vary from 0 to 2, and the lower the score, the higher the quality of the forecast.

In order to get the score for the entire dataset we take the average

$$B = \sum_{i=1}^{N} B_i$$

where N is the number of cases. Clearly, the total quality of a model expressed with the Brier score ranges from 0 to 2. The lower the score the better the quality of the prediction.

2.6.4.3. Average logarithmic loss score

The average logarithmic loss score is calculated with the expression

$$L = \frac{1}{N} \sum_{i=1}^{N} -\log(p_i),$$

where p_i is the probability predicted for the correct state in case i, and N is the number of cases whose class variable is to be predicted [Morgan & Henrion, 1990]. Average logarithmic loss values are calculated using the natural logarithm, and are between 0 and infinity inclusive, with the value equal zero indicating the best performance.

2.6.5. Other validation methods

Another class of validation methods are measures that quantify the distance between joint probability distributions. One of the distributions is the joint probability distribution encoded by the network under validation, whereas the other one is some "true" or "target" probability distribution. They include Euclidean distance, spherical pay-off, cross-entropy, also known as Kullback-Leibler divergence, Cosine distance, and Jensen-Shannon accuracy [see Bang *et al.*, 2003; Heckerman *et al.*, 1995; Heckerman and Nathwani, 1992].

Another useful measure of the quality of structural learning is the *structural difference* measure. This measure is intended to assess the degree to which the learned network has captured causal relationships. For instance, Heckerman *et al.*

[1994] calculate for each node the symmetric difference of the parents of the node in the gold-standard network and the parents of this node in the learned network. Then, the structural difference between the networks is given as the summation over each node.

Another class of measures are the *information criteria*, such as Bayesian information criterion (BIC) [Schwarz, 1978], and Akaike information criterion [Akaike, 1974]. The basic idea of these measures is to select the network structure that best fits the data, and at the same penalized for the number of parameters that need to be specified to define the joint probability distribution associated with the network.

2.7. Use

The following types of probabilistic inference are possible in the Bayesian network framework:

- Diagnostic (backward, bottom up),
- Causal (forward, top down),
- Explaining away (inter-causal),
- Combination of the above.

In short, the diagnostic inference refers to inferring beliefs about causes on the basis of effects, or symptoms. Inferring in the other direction, i.e., from causes to effects, can be called causal. An interesting type of inference is the inter-causal inference, in which on the basis of the evidence on one of possible causes we can update our beliefs about other causes. We explain and illustrate these capabilities of Bayesian network modelling in the context of the CS&L research in Chapter 4.

In consequence, in the Bayesian network classification there are no pre-specified class variables, or, in other words any variable can be treated as a class variable in one and the same Bayesian network model.

Furthermore, an advantage is that when no data exist on some variables we can continue inference (prediction), so that no special treatment, such as adding another state "missing" to a variable is required. The treatment of missing data in prediction tasks is therefore by its nature very economic.

2.7.1. Probabilistic inference

Joint probability distribution over a domain, encoded with a BN, can be queried for any probability of interest just as if we had a joint probability table for the domain and acted on this table by summing probabilities in relevant cells. The queries can be classified into the following categories: single marginal, subjoint marginal, all marginals, arbitrary set of queries, conditional (single, subjoint, and total) marginals, Boolean queries, most probable explanation, and maximum a posteriori [D'Ambrosio, 1999]. We describe the possible queries in the following subsections. Let us assume that the example Bayesian network in a domain X includes the variables $X_1, X_2, \ldots, X_n$.

- Single marginal

The most obvious type of query is the probability distribution over the states of a single random variable prior to any evidence. Formally, this probability could be presented as:

$$P(X_k \mid S, \Theta) = \sum_{i \neq k}^{n} P(X_i \mid S, \Theta),$$

where X_k is a variable of interest, S is the structure and Θ are parameters.

- Subjoint marginal

It is very easy to retrieve a marginal probability distribution for any set of variables. The marginal of a subset of random variables X_i and X_j can be shown as

$$P(X_i, X_j \mid S, \Theta) = \sum_{k \neq i,j}^{n} P(X_k \mid S, \Theta)$$

- All marginals

Extending the subjoint marginal query to other remaining variables we can retrieve the probability related to the case in which all variables in the model are instantiated with a particular state.

$$P(X_1, X_2, \ldots, X_n \mid S, \Theta) = \sum_{i}^{n} P(X_i \mid S, \Theta)$$

- Arbitrary set of queries

The next query is a natural generalization of the all-marginals query, in which we ask for an arbitrary subset of subjoints.

- Conditional (single, subjoint, and total) marginals

Another very important query is the conditional, which covers the situation when we need a marginal probability given some evidence.

- Boolean queries

We could also make a Boolean query of the type $P((X_1 = x_1 \wedge X_2 = x_2) \vee (X_1 = x_1 \wedge X_4 = x_4) \mid S, \Theta)$.

- Most probable explanation

The most probable explanation query is similar to the all-marginals mode in that it returns the configuration $X_1, X_2, \ldots, X_n$ with the most probability. In other words, in a table with a joint probability distribution, it would be a cell with the highest probability. Clearly, we can also acquire the most probable configuration after an evidence is entered. A variant of this type of query is to ask for m most likely configurations, instead of a single most likely one. In Hugin, one of the most popular Bayesian network software environments [Andersen *et al.*, 1989], this query is very easy to realize in practice using the so-called *max-normal* propagation, which is a by-product of the typically used sum-normal propagation [Spiegelhalter and Lauritzen, 1990; Dawid, 1992].

- Maximum a posteriori probability

In the case of most probable explanation we are interested in the highest probability in the full joint probability table $P(X_1, X_2, \ldots, X_n)$. Often, we are however interested in the most probable configuration over a subjoint, for example $P(X_1, X_2)$ with the highest value. This query boils down to finding the cell in the subjoint probability table $P(X_1, X_2)$ having the highest probability. Again, the most likely instantiation can be in the prior or with respect to evidence.

2.7.2. Inference algorithms

It is the existence of the inference algorithms that make the use of Bayesian networks so efficient and attractive to a modeller. We must keep in mind however, that in the general case, doing inference in Bayesian networks is an NP-hard problem [Cooper, 1988]. There are many algorithms in the literature that can be categorised as exact, approximated and symbolic [Castillo *et al.*, 1995].

The *exact* algorithms of inference algorithms have as their output probabilities that are results of simple arithmetic operations, such as multiplication, addition, etc. The resultant probability of interest is thus mathematically exact value. The most efficient methods include factoring [Darwiche, 2003], bucket elimination [Dechter, 1996], and clustering algorithm, also known as the junction tree or click tree algorithm [Lauritzen and Spiegelhalter, 1988; Jensen *et al.*, 1990]. The junction tree algorithm is regarded as the most computationally effective algorithm in general, but for models with a specific dependency structure it can turn out to be intractable. The basic idea is to cluster nodes together to obtain a singly-connected structure, and then execute a message passing scheme along that structure. First, we create a moralized graph from the original Bayesian network by "marrying" parents, i.e., connecting parents of each node, if the parents are not connected in the original Bayesian network. Next, we triangulate a moralized graph is, so that it contains no cycles of more than three nodes; by doing this we obtain a triangulated graph. Subsequently, a tree of cliques, also called join tree or junction tree, is created from the triangulated graph. Then, it computes probabilities for the cliques during message propagation and the individual node probabilities are calculated from the probabilities of cliques.

In contrast to exact methods, the methods of *approximated* probability updating retrieve only rough approximations of the conditional probabilities. These methods are based on the idea of stochastic simulation and take into account the causal ordering of variables encoded by a given model. For instance, if we are interested in prior marginal distribution of a variable of interest, we first simulate a state of the root nodes using a pseudorandom generator according to the prior distribution of these nodes. Next, given the sampled values of the parents, we sample the state of child nodes accordingly, and we continue following the flow of influence in the model. Then, we sample the next case in the same way. If we sample long enough, the probability distribution of the variable of interest should became stable and converge, so that we can take this distribution as approximation of the marginal probability of interest. This procedure is known as logic sampling [Henrion, 1988].

Algorithms of *symbolic propagation* allow for updating in Bayesian networks in which some probabilities are given not numerically but symbolically as parameters. As a result, they compute the probabilities of interest in symbolic form as a function of the symbolic parameters [Castillo *et al.*, 1995].

2.7.3. Explanation

There is an extensive body of research generally concerning explanation in the Bayesian network literature. However, virtually the whole body of this research is approached from the point of view of expert systems. To be more precise, it is argued that an important characteristic of decision systems, and Bayesian networks in particular, is explanation how the system has reached its conclusions. In this context, methods have been proposed that should explain which evidence is most important for a given conclusion, or how the evidence has traversed the network [e.g., Madigan *et al.*, 1997].

A part of the explanation is sensitivity analysis [Jensen, 2001]. Sensitivity analysis is one of the topics in Bayesian network modelling that we exploit heavily in this dissertation, especially in Chapter 6. Therefore, we refrain with the discussion of this area until Chapter 6.

2.8. Bayesian networks in the marketing literature

Based on the search in the marketing literature, we have found several publications whose main objective was to apply Bayesian networks in specific business and/or marketing problems [e.g., Shenoy and Shenoy, 1999; Anderson and Lenz, 2001; Alexander, 2000; Blodgett and Anderson, 2000]. Of these articles only two deal with problems in marketing research. Blodgett and Anderson [2000] modelled consumer complaint process for explanation and prediction of consumer behaviour after experiencing dissatisfaction with a product. Anderson and Lenz [2001] apply Bayesian networks to model and assess the impact of potential changes in the decision-making process of a large global manufacturing organisation. The remaining two articles Shenoy and Shenoy [1999] make use of

Bayesian network models in the financial analysis. All these articles are rather application-specific and none of the issues required from the point of view of development and justification of CS&L theory, as discussed in Section 1.2, have been addressed in these publications.

2.9. Conclusions

In this chapter, we presented a review of key issues relating to the history, definition and the fundamental use of Bayesian networks. This review was based on the existing data mining/machine learning literature. We have also mentioned a few articles that appeared to date in the marketing literature.

We have noticed that a lot of issues central for the proliferation and successful use of Bayesian networks have been addressed neither in the machine learning literature nor in the marketing literature. In view of this review, we must conclude that many fundamental issues in the use of Bayesian networks still remain open and are subject to further development, especially in the context of special requirements in theoretical and practical Customer Satisfaction and Loyalty research, as discussed in Section 1.2.

3. Current research on Customer Satisfaction and Loyalty

In the first part of this chapter, we present a short literature overview on selected topics in Customer Satisfaction and Loyalty research. We would like to stress that this overview is not intended to be complete or representative of the entire domain of CS&L research. Since it is not the main aim of the thesis to get insight into theoretical relationships between concepts involved in the CS&L research, the objective of this review is rather to introduce the most important literature and highlight only some concepts as far as they are relevant for the case studies that we consider further. To be more precise, we discuss only those concepts that appear in the data of the case studies.

In the second part of this chapter, we present an overview of the dominant data analysis and techniques applied today in Customer satisfaction and Loyalty research. The presented techniques are used to corroborate scientific hypotheses and to develop theories. We do not aim to provide a detailed account of evaluation of these techniques, but only characterize them in terms of the main features.

The chapter is organized in the following way. In Section 3.1, we give an overview of the current state of knowledge and theoretical underpinnings in CS&L research. Then, in Section 3.2, we provide a short exposition to the modelling techniques applied today in CS&L research.

3.1. Customer Satisfaction and Loyalty literature

In this section, we will discuss the concepts of perceived service quality, image, satisfaction, trust, commitment and involvement. This is not to maintain that these psychological constructs comprise an exhaustive list of causal variables of customer loyalty. As a matter of fact, many researchers and practitioners emphasize the importance of others, e.g., perceived value.[1] In Chen *et al.* [2003], perceived value of an e-store is, based on Zeithaml [1988], defined as "the overall assessment of the utility of a product based on what is received and what is given." The authors proposed three value's components: value-for-money, trust, and shopping efficiency, and found support for hypotheses that these value components would be positively related to e-store loyalty intentions, although value-for-money showed the least effect. These constructs however are believed to "summarize consumers' knowledge and experiences with a particular firm and guide subsequent actions of the customer" [Garbarino and Johnson, 1999]. Furthermore, we discuss only those concepts that appear in the data of the case studies.

[1] We note that customer value is a concept strongly correlated with customer satisfaction [Van Riel *et al.*, 2001]

Recently, more and more interest in the marketing community is attached to customer loyalty in the context of web and online services. It is known that customer retention in e-services is of paramount importance [Reichheld and Schefter, 2000], because 1) customers can quite easily evaluate and compare the benefits of competing e-services, and 2) switching costs are low [Van Riel *et al.*, 2001. However this field is still a neglected area in the marketing research [Liljander *et al.*, 2001], and many constructs lack clear conceptual and operational definitions [e.g., Aladwani and Palvia, 2002]. So, generally, the field of customer e-loyalty seems still to be in its infancy.

Electronic commerce is able, as never before, to make that the promise expressed by Dwyer *et al.* [1987] come true: it offers improved conduit of communication from customers, and it allows marketers to obtain high quality information. But is there any need to consider e-loyalty apart from traditional customer loyalty? It is a common belief that with reference to online loyalty "the same rules apply". However, in our opinion, it is important to note that most of research on customer evaluations and quality of services has been conducted with respect to services that are characterised by personal interactions between consumers and employees [Van Riel *et al.*, 2001].

Following Bansal *et al.* [2003] it is believed that some of the basic premises, such as the quality-satisfaction-loyalty chain, that underlie offline relationships apply to online environments, still there are also likely to be fundamental differences [e.g., Ranaweera et al., 2004].

Each construct will be described in terms of its conceptual as well as operational definitions. According to Jacoby and Chestnut [1978], conceptual definitions encompass "the essence of what we mean when we speak about a particular item, phenomenon, or event". Next, operational definitions, i.e. detailed descriptions of the procedures used to measure the concept are given. Then, for each construct also theoretical constructs that are likely to have causal effects on them will be mentioned, and specific activities that marketing managers are basically in control of and can apply in order to develop desired effects of construct's arousal will be covered. Subsequently, the construct's behavioural as well as attitudinal consequences are explicated. Lastly, existing literature on conceptualisation of construct relevant for the online context is provided.

We begin this review with the notion of customer satisfaction. Next, we proceed with discussing customer loyalty, and other theoretical constructs.

3.1.1. Customer Satisfaction

Satisfaction is an issue that has received considerable attention by academics as well as practitioners and is well documented in the literature, although, as Peterson and Wilson [1992] stated, studies of customer satisfaction are best characterized by lack of definitional and methodological standardization. It has been considered as a central construct in marketing literature [Erevelles and

Leavitt, 1992]. However, its relation to loyalty is not well-specified and still remains to be investigated [Oliver, 1999].

Customer satisfaction is accepted as a critical concept in marketing thought and consumer research [e.g., Peter and Olson, 1996], even though some researchers regard this notion as confounding and opt for a conceptually similar concept of perceived service quality.

Oliver [1999] argues that satisfaction is a necessary step in loyalty formation but becomes less significant when other mechanisms, such as social bonds or personal determinism come into play.

1) Conceptualisation

There is a plethora of customer satisfaction definitions in the marketing literature. It has been defined as "an evaluation of the perceived discrepancy between prior expectations and the actual performance of the product as perceived after its consumption" [Tse and Wilton, 1988], or as "a global evaluative judgment about product usage/consumption" [Westbrook, 1987].

Oliver [1981] stated that "satisfaction is a summary psychological state resulting when the emotion surrounding disconfirmed expectations is coupled with the consumer's prior feelings about the consumption experience". Satisfaction is a function of prior expectations, post-purchase evaluations of product (service) performance, and level and nature of disconfirmation.

In general, Giese and Cote [2000] found that three overall components within virtually every definition of satisfaction might be identified that capture the specifics of the concept. These components are: 1) a response (affective or cognitive), 2) the response concerns a particular focus (e.g. expectations, product, and consumption experience), 3) the response takes place at a particular point in time (e.g. after choice, after transaction, after consumption, based on accumulated experience).

Cote *et al.* [1989] argue that none of the conceptualisations pertaining to satisfaction as being perceived in some point in time is appropriate because satisfaction can vary dramatically over time. They suggest that the satisfaction can be realized by customers only at the time while an explicit evaluation occurs, for instance, when consumer is being surveyed. Also Schwarz and Strack (as cited in McClendon and O'Brien, 1988] hypothesized that survey participants often have not thought about a study topic until they are asked a specific question. Peterson and Wilson [1992] speculate that the answers to several types of questions may well be related to survey participant's mood at the time of responding.

There exists a discrepancy as to the nature of satisfaction in terms of its cumulativeness vs. single encounter specific. Some authors lean towards conceptualization as being cumulative, others as being transaction specific. Satisfaction that occurs strictly at time of the service delivery is referred to as service encounter satisfaction [Bitner, 1990; Parasuraman *et al.*, 1988], whereas overall customer satisfaction is relationship-specific, that is, it is the cumulative effect of a set of discrete service transactions or encounters with the service provider over a period of time [Bitner and Hubbert, 1994; Oliver, 1997; Rust and Oliver, 1994]. The two types are related but may have different factors tat influence them. Westbrook [1981] shows that satisfaction with a retail store is an accumulation of separate satisfaction evaluations with the salesperson, store environment, products and other factors.

Historically, the earliest attempts to capture the phenomenon of customer satisfaction were directed at a conceptual model, which postulated a direct causal link between the performance of product/service attributes and overall state of satisfaction [Oliver, 1997]. According to this representation, there is actually no intermediate psychological state, nor cognitive process that mediates the formation of (dis)satisfaction judgments. The approach can thus be summarized as "a black-box" model of customer satisfaction [*idem*], because consumer thought processes are not taken account of as a part of this phenomenon. This approach however has been questioned by most scholars, and is rather neglected in today's advanced customer satisfaction research as it is missing good theoretical groundings. Nevertheless, it still remains applied by many companies in traditional attribute performance analysis [Naumann and Giel, 1995; Oliver, 1997].

The primary thread of debate in the satisfaction literature nowadays is focused on the nature of the cognitive and affective processes that result in the consumer's state of mind referenced to as satisfaction. In line with this stream of research, the two dominant approaches compete whether satisfaction can be best described as an evaluation process [e.g. Fornell, 1992; Oliver, 1981; Yi, 1990] or as an outcome of an evaluation process [Tse and Wilton, 1988].

With regard to the view of satisfaction as an outcome of an evaluation process, customer satisfaction is viewed as a state of fulfilment that is associated with reinforcement and arousal. In the "satisfaction-as-states" framework developed by Oliver [1989], several types of satisfaction have been identified as a potential state of fulfilment, including: "satisfaction-as-pleasure", "satisfaction-as-relief", "satisfaction-as-novelty", "satisfaction-as-surprise", "satisfaction-as-contentment". In line with this paradigm, satisfaction is defined as "a pleasurable level of consumption-related fulfilment" [Oliver, 1997].

The second, and according to Oliver [1999], more prevailing mainstream of research on Customer Satisfaction/Dissatisfaction (CS/D) as an evaluation process is based on the paradigm of expectations' disconfirmation [Churchill and Suprenant, 1982; Oliver, 1980]. Its central assumption is that consumers form prior expectations (e.g., caused by commercials, advertisements, experience, etc.) towards product/service performance, which later serve as standards against which actual product/service performance is evaluated [Oliver, 1980; Churchill and Suprenant, 1982]. A comparison of expectations and actual perceived performance results either in confirmation or disconfirmation. In case prior expectations are exactly met, a mere confirmation takes place. Otherwise, disconfirmation occurs, i.e. the perception of a discrepancy between performance and expectations. Within disconfirmation, two types, positive and negative, may be identified. Positive disconfirmation occurs when perceptions exceed expectations and negative disconfirmation occurs when expectations exceed perceptions. According to this paradigm, satisfaction is the result of positive disconfirmation and simple confirmation, whereas negative disconfirmation leads to dissatisfaction. Moreover, it is also believed that expectations have an indirect influence on satisfaction via disconfirmation, whereas performance can have both an indirect via disconfirmation, as well as direct effect on (dis)satisfaction. The two different types of conceptualisations may be jointly applied to a particular context enhancing thus predictive power of satisfaction as a measure related to loyalty [Rust and Oliver, 1994].

The application of process definitions is regarded relevant for brief service encounters as well as for services that are delivered or consumed over a certain period of time [Oliver, 1996; de Ruyter and Bloemer, 1998]. Oliver [1996] argues that this is a typical aspect of service satisfaction. However, the two different types of conceptualisations may be jointly applied to a particular context enhancing thus predictive power of satisfaction as a measure related to loyalty [Rust and Oliver, 1994].

Bloemer and Kasper [1995] argue to distinguish between latent and manifest satisfaction. They conclude that the moderator effect of consumer's elaboration upon the brand choice decision, operationalized as the degree of involvement and deliberation within the purchase exists. Consequently, increase in customer's manifest satisfaction, understood as high levels of motivation and capacity of evaluation a brand choice (as relating to expectations and performance evaluation) has a larger effect on true brand loyalty than the same increase in latent satisfaction.

Geyskens and Steenkamp [2000] distinguish between another two types of satisfaction: economic, and social satisfaction. Their study concerns explicitly business channel members' satisfaction. In their conceptualisation, economic satisfaction is defined as a channel member's evaluation of the economic outcomes that flow from the relationship with its partner, whereas social satisfaction depicts a channel member's evaluation of the personal contacts and interactions with its exchange partner.

2) Operationalization

In practical CS/D measurement studies, it is however approved to measure satisfaction directly [Naumann and Giel 1995], therefore we assume the traditional, non-mediated model of satisfaction, allowing thus for direct links from product/service attributes' performance to (dis)satisfaction.

Sample operationalizations of satisfaction conceptualized as a process include items like: "this service is in agreement with my expectations." Measures of outcome-related definitions contain items like: "I am satisfied with the service", or they explicitly pertain to the level of customer's fulfilment of satisfaction as a state.

Cronin and Taylor [1992] demonstrate that one item self-report measure of satisfaction has a stronger significant effect on purchase intentions than 22-item operationalization of service quality.

3) Antecedents

In line with the expectancy disconfirmation model, customer satisfaction has three antecedents: perceived performance, expectations, and disconfirmation of expectations. Besides these constructs, research on customer satisfaction has focused also on modelling the effects of affect and equity on buyers' level of satisfaction [Szymanski and Henard, 2001].

There is an agreement that expectations as well as performance are not formed on an aggregate level but for each product/service attribute separately. Various researchers have approached expectations differently. Traditionally, the role of expectations has been modelled in one of two ways. One is the role as anticipation; the other one is the role as comparative referents [Szymanski and Henard, 2001]. With regard to the role as anticipation, expectations can be formed in terms of ideal product performance, minimal expectations, "will" expectations, and other types. Currently, expected product performance defined as a product's most likely performance ("predictive performance") is the most common presumption used in CS research. Some models use equity expectations based on what the consumer believes reasonably should occur given the product/service price [Oliver and Swan, 1989].

There are a lot of empirical studies that confirm the influence of attribute level performance [Bearden and Teel, 1983; Bolton and Drew, 1991; Mittal *et al.*, 1998; Oliva et al., 1992; Oliver, 1993; Spreng *et al.*, 1996]. Performance is not only antecedent through disconfirmation, but also indirectly [Churchill and Suprenant, 1982; Oliver and De Sarbo, 1988]. The impact of attribute-level performance could be positive or negative depending on the attribute.

As noted earlier, confirmation/disconfirmation of expectations is another antecedent concept for satisfaction [Patterson *et al.*, 1997; Anderson and Sullivan, 1993; Churchill and Suprenant, 1982; Tse and Wilton, 1988, Oliver, 1980]. Moderate satisfaction is a result of confirmed expectations, positively disconfirmed standards lead to high satisfaction, and negatively disconfirmed expectations lead dissatisfaction. Expectations play thus the ole of comparison standards in this respect [Oliver, 1997; Erevelles and Leavitt, 1992].

Besides the performance, also the prior experience [Bolton and Drew, 1991; Cadotte *et al.*, 1987; Vredenburg and Wee, 1986]. Woodroof *et al.* [1983] argue that favourable prior experience with a service provider increases the likelihood of the favourable evaluation of the current service encounter as well as the overall evaluation of the service provider. Vredenburg and Wee [1986] found that favourable prior experience resulted in higher satisfaction levels.

Another antecedent of satisfaction is image [Andreassen and Lindestad, 1998].

4) Consequences

The most often analysed outcomes of satisfaction include complaining behaviour, negative word of mouth, and repurchase intentions [Szymanski and Henard 2001]. Consumers tend to complain to sellers to relieve cognitive dissonance when the consumption experience is dissatisfying [Oliver, 1987], especially when the problem leading to dissatisfaction is severe, the degree of external attribution of blame is to the retailer or manufacturer, or the likelihood to redress is high [Folkes, 1985; Blodgett and Anderson, 2000].

Another form of complaining behaviour is negative word-of-mouth behaviour to other consumers that increases in the face of dissatisfying experience [Nyer, 1999].

With respect to repurchase intentions, Garbarino and Johnson [1999] demonstrate that depending on the level of customer's relationship orientation, satisfaction is either a mediating construct between trust and commitment on the one hand, and future behavioural intentions on the other (for low relational customers) or it does not relate to future intentions (for high relational customers).

Oliva *et al.* [1992] found that the link between satisfaction and loyalty is of non-linear nature. Furthermore, they argue that there exists a certain threshold of satisfaction, above which loyalty will rapidly increase, but on the other hand, below which loyalty remains unaffected over a range of satisfaction levels. The link between satisfaction and loyalty is amplified by experiencing of positive emotions during the service delivery process for highly involving services [Bloemer and de Ruyter, 1999]. The recent advances in the satisfaction research [e.g., Cronin *et al.*, 2000] have proved that satisfaction as well as service quality and perceived value are directly linked to behavioural intentions, but additionally, indirect effects of the service quality via satisfaction enhanced their impact on behavioural intentions.

According to Smith and Rutigliano [2003], customer satisfaction is "merely the entry point for achieving a deeper foundation that rests on total customer engagement."

In general, a variety of studies have found evidence that higher satisfaction leads to greater customer loyalty [Anderson and Sullivan, 1997; Bolton and Drew, 1991; Oliver and Swan, 1989; Fornell, 1992].

Based on the transaction-driven nature of satisfaction, several writers claim that the cumulative aspect of customer satisfaction has effect on image [Bolton and Drew, 1991; Oliver and Linda, 1981; Fornell, 1992].

5) e-Satisfaction

There is small but growing body of research on conceptualisation and drivers of e-satisfaction [Bansal *et al.*, 2004]. Some of the studies that appeared to date are listed in Table 2.1.6.

Anderson and Srinivasan [2003] define e-satisfaction as "the contentment of the customer with respect to his or her prior purchasing experience with a given electronic commerce firm." The authors found support for the link between e-satisfaction and e-loyalty, but on the customer level this link would be accentuated by convenience motivation, and purchase site, and suppressed by inertia, whereas on the business level trust and perceived value significantly accentuate this link.

Antecedents

As we have mentioned, relatively few studies have examined the factors that make e-customers satisfied with their online experience. One of the first studies reported in the academic literature is the study of e-shopping by Szymanski and Hise [2000]. In their investigation of a conceptual model of e-satisfaction they found that convenience, product information, site design, and financial security are the most important factors driving online-specific satisfaction. Van Riel et al. [2001] found that e-satisfaction is influenced by service dimensions as satisfaction with core service, supplementary services, and user interface. Furthermore, e-satisfaction strongly influences intention to return to web site, but has no direct influence on perceived value.

Ranaweera *et al.* [2003] posit that website characteristics such as ease of use, web content, security/privacy, interactivity, reliability, customer service, and price are drivers of transaction-specific satisfaction with a website.

In the study by Bansal *et al.* [2003], support was found for the proposition that web site characteristics, such as ease of use, product selection, information availability, price were the major driver of overall web site satisfaction, while customer service played a significant but lesser role. Furthermore, they found a linkage between website characteristics and stickiness.

Simon [2001] studied the impact of culture and gender on satisfaction with websites. He found a difference between perceptions of males and females in that females within certain cultures have widely different preferences from their male counterparts regarding web site attributes.

Consequences

In a study across product categories and world regions, Lynch *et al.* [2001] have found that customer affect, conceptualised as experience of feelings of happiness, excitement, and enthusiasm, was significant in three of the twelve regressions predicting purchase and loyalty.

Anderson and Srinivasan [2003] found that the link between e-satisfaction and e-loyalty is positively moderated by convenience motivation, and purchase site, and negatively by inertia.

Interesting results are delivered by Shankar *et al.* [2003], who show that whereas the levels of customer satisfaction for a service chosen online is the same as when it is chosen offline, loyalty to the service provider is higher when the service is chosen online than offline. The relationships between satisfaction and loyalty is thus stronger online than offline.

3.1.2. Customer Loyalty

The first academic investigations related to the subject of brand loyalty can be traced back to the year 1923, beginning with the work by Copeland [1923]. This early work in the field lacked, however, both well-underlying conceptual and methodological basis. The major developments towards the specification of the proper research methodology in the area were worked out in the late 1960s, and the 1970s [e.g., Day, 1969; Jacoby, 1971; Cunningham, 1956].

At the core of the customer loyalty studies lies the concept of repeat purchase behaviour, which can be regarded as some degree of repetitive purchase of the same brand by the same buyer. There are in general two approaches that pertain to the nature of such a construct [Jacoby and Chestnut, 1978]. At the core of the first approach is the suggestion of a strong random component that underlies basic changes in the market structure. This view, regarded as *stochastic*, assumes that even if repeat purchasing is caused by some variables, their multiplicity and complexity is so immense that it makes the purchasing behaviour virtually an unpredictive concept and thus authorizes to claim that it is a stochastic process [e.g., Ehrenberg, 1972].

In contrast with this trend is the second philosophy of *deterministic* nature of repeat purchasing behaviour. In line with this approach there is a limited number of causes that directly influence the repeated purchasing. Those causes can be isolated from each other and then stimulated by the marketing manager in order to bring about the desired effects of repeat patronage. The customer loyalty research is based on this approach and thereby is focused on repeat purchase behaviour that can be reasonably explained by means of some underlying constructs, as beliefs, attitudes or opinions. In this sense, deterministic orientation can be applied to a subset of repeat purchase behaviour, which can be termed as brand loyalty. The first authors that provided conceptually clear and precise definition of brand loyalty were Jacoby and Chestnut [1978]. Their definition is expressed by a set of six necessary and collectively sufficient conditions as "(1) the biased (i.e., nonrandom), (2) behavioural response (i.e., purchase), (3) expressed over time, (4) by some decision-making unit, (5) with respect to one or more alternative brands out of a set of such brands, and (6) is a function of psychological (decision-making, evaluative) processes". Nevertheless, the conceptual framework for the definition of customer loyalty provided by

Jacoby and Chestnut [1978] seems to be most widely accepted among marketing researchers. It stimulated the vast body of work in the field.

Oliver [1999] describes loyalty as "a deeply held commitment to re-buy or repatronize a preferred product/service consistently in the future, thereby causing repetitive same-brand or same brand-set purchasing, despite situational influences and marketing efforts having a potential to cause switching behaviour".

Dick and Basu [1994] define customer loyalty in a way that is conceptually similar as the one of Jacoby and Chestnut [1978] denoting loyalty as "the relationship between relative attitude and repeat patronage". They argue to consider a relative attitude as opposed to an absolute one due to potential variations in strength and differentiation of people's attitudes with respect to other targets. Such a definition has a number of advantages. First of all, it enables to avoid the mistake of treating loyalty as a behaviour only resulting in repeat purchasing. It has been evidenced many times that repeat purchasing is not always a result of loyalty, as the customer may simply be forced to buy a particular brand because of some situational factors, such as shelf positioning, or because of subjective norms. Subjective norms, also referred to as social norms, are one's beliefs that they should, or should not, do something caused by their vulnerability on influence of others' beliefs. They define this case of low attitude accompanied by high repeat patronage as "spurious loyalty" [Dick and Basu, 1994]. Some authors call this kind of loyalty as "driven by inertia" [Assael, 1992]. On the other hand, high relative attitude with low repeat purchasing is defined as "latent loyalty." Dick and Basu [1994] suggest that also in this case "loyalty" occurs due to marketplace environment, in which subjective norms or situational effects are stronger than the relative attitude.

2) Operationalization

Good measures of loyalty should capture attitudinal as well as behavioural factors of loyalty; however, in the past the marketing literature was abundant in the behavioural measures of loyalty. Jacoby and Chestnut [1978], for instance, cited 53 definitions, a great magnitude of which were operational and devoid of theoretical meaning.

Following Jacoby and Chestnut [1978], Dick and Basu [1994] strongly suggest to operationalise loyalty as the index of the strength of the relation between the attitude and repeat purchasing. Based on slight modification of this paradigm, Bloemer and Kasper [1995], for instance, measure loyalty as an outcome of multiplication of the score for customer commitment times the score for future purchase intentions.

Though researchers agree that incorporating measures for future purchase intentions, instead of directly measuring purchase behaviour is rather tentative measure of customer loyalty, as follow-up studies that might verify these intentions are rarely performed [e.g., Oliva *et al.* 1992], most studies apply this measure and not the other. This limitation seems to result from practical reasons, and not methodological ones.

3) Antecedents
All constructs discussed in this review are regarded in the CS&L literature as antecedents of customer loyalty.

4) e-Loyalty
We have found very little studies that clearly define the concept of e-loyalty. Some studies may provide however valuable indications. One of these studies is by Anderson and Srinivasan [2003], in which e-loyalty is defined as "the customer's favourable attitude toward an electronic business resulting in repeat buying behaviour." Gommans *et al.* [2001] conceptualise e-loyalty and discuss similarities and differences between traditional brand loyalty and e-loyalty.

The electronic nature of the service encounters between customers and companies enables easy and efficient measurement of actual e-loyalty behaviour. Behavioural outcomes of e-loyalty would include web site usage-related measures, such as visit frequency, visit duration, visit scope, visit focus, or stickiness [Cutler and Sterne, 2000]. Stickiness has been used in a recent study [Bansal *et al.*, 2003] as a measure of actual browsing behaviour. These outcomes can have also much utility value, as predominantly they are related directly to company profits, e.g., through banner views, and advertisements, but they can indicate customer's level of online loyalty. Moreover, measurement of behavioural e-loyalty understood as usage of web services in light of alternatives, similar to the "share-of-wallet" principle, can be easily achieved.

Authors	Dependent Variable	Drivers	Comments
Srinivasan, et al., 2002	e-loyalty	Customization, Contact Interaction, Cultivation, Care, Community, Choice Character	A key feature of the study was that it relied on data collected from a survey of online consumers. All data collected were from the same instrument. Surprisingly, convenience (i.e. ease of use) was not a driver. E-loyalty was significantly related to word of mouth and willingness to pay more.
Chen et al. 2003	e-store loyalty	Perceived value (value-for-money, trust, and shopping efficiency)	
Anderson and Srinivasan, 2003	e-loyalty	E-satisfaction	The link between e-satisfaction and e-loyalty is moderated by convenience motivation (+), purchase site (+), and by inertia (-).
Gommens et al., 2001	e-loyalty	Website and technology, customer service, value proposition, trust and security, brand building	
Zeithaml et al., 2000	e-service quality	- Perceived Convenience (access, ease of navigation, efficiency,	Navigation, efficiency and access are "new" to the online environment and capture aspects of ease of use and

		flexibility) - Perceived Control (reliability, personalization, security/privacy)	usefulness which underlie acceptance of information technologies (Davis, 1989) Paper was conceptual
Zeithaml et al., 2002	e-SQ	Technology readiness, Demo/sociographics	
Francis and White, 2001	PIRQUAL (Perceived Internet Retail Quality Model)	Web store functionality, Product attribute description, Ownership conditions, Delivered products, Customer Service, Security	Except for product attribution description, the factors were significantly related to future visit and purchase intentions. Results confirmed proposition that online satisfaction was a function of the purchase experience, the delivery experience and the customer service (e.g. responsiveness, fix problems) / security experience
Donthu, 2001	SITEQUAL (Internet Shopping Quality)	- Site-related factors (ease of use, - aesthetic design, processing speed, security) - Vendor-related factors (competitive value, clarity of ordering, corporate and brand equity, product uniqueness, product quality assurance)	A validation study indicated that SITEQUAL was directly correlated to shopping likelihood, attitude and loyalty.
Loiacono et al., 2002	WebQual? (Web site quality – predicts Web site reuse)	- Ease of use (ease of understanding, intuitive operations) - Usefulness (informational fit-to-task, interactivity, trust, response time) - Entertainment (visual appeal, innovativeness, flow emotional appeal) - Complementary Relationship (consistent image, online completeness, better than other channels)	The model exhibited both reliability and validity, and correlations between the composite WebQual measure and intention to purchase and intention to revisit were significant.
Song and Zinkhan, 2003	Web site quality	Site design (user interface, information access, fulfilment policy)	Web site quality will positively affect consumers' purchase, repeat purchase, and loyalty
Ranaweera et al. 2004	Satisfaction with a website	Website characteristics (ease of use, web content, security/ privacy, interactivity, reliability, customer service, price)	
Szymanski and Hise, 2000	e-satisfaction	Convenience, Site design, Financial security, Product information.	Compares e-tail to retailing satisfaction, based on Web site characteristics and did not include potential drivers such as customer service.
Van Riel et al., 2001	e-satisfaction	Satisfaction with core service, Supplementary services, User interface	e-Satisfaction strongly influences e-loyalty, but has no direct influence on perceived value
Bansal et al., 2003	Web site satisfaction	Web site characteristics (ease of use, product selection, information availability, price)	Customer service played a significant but lesser role

Hoffman et al., 1999	Trust		Two dimensions of online trust: environmental control, and secondary use of information control

Table 3.1.1 Drivers of e-loyalty (partly from Bansal et al. [2003]).

Table 3.1.1 presents a short overview of selected recent studies in the academic literature on drivers of customer e-loyalty. As far the outcomes of e-loyalty are concerned, the traditional outcomes of customer loyalty, such as recommendations, complaint behaviour, purchase intentions, repatronage decisions, search motivation, etc., are also adequate.

3.1.3. Service Quality

Perceived service quality is considered an essential determinant of success and survival in today's competitive environment [Dawkins and Reichheld, 1990]. It is widely recognized to have a strong effect on behavioural intentions towards service providers [e.g., Parasuraman *et al.*, 1994].

Multiple studies have taken an effort to determine what is the exact relation between service quality and loyalty as well as what aspects of service quality are considered by customers when evaluating service performance.

The need for a special construct to assess service quality stems from the specific nature of a service construct, which can be characterized, in contrast to goods, by its three unique features: intangibility, heterogeneity (their performance varies across deliveries, time, customer), and inseparability of production and consumption. Due to these features and, consequently, absence of objective measures, the most appropriate method to evaluate the service quality is to evaluate consumers' perceptions.

1) Conceptualisation

In the first attempt by Parasuraman *et al.* [1985] to analyse the construct of service quality the essential theme was that "service quality perceptions result from a comparison of consumer expectations with actual service performance". Their conceptualisation was strongly related to one for satisfaction in that it pertained directly to the disconfirmation paradigm. Furthermore, they proposed to describe the level of perceived service quality as defined, i.e. the discrepancy between expectations and perceived performance, in terms of satisfaction (e.g., when expectations are met with performance, service quality is perceived as satisfactory). The focus group interviews resulted in discovering 10 common for most types of service criteria, along which consumers assess perceived service quality, namely: reliability, responsiveness, competence, access, courtesy, communication, creditability, security, understanding/knowing the customer and tangibles.

The second milestone study by Parasuraman *et al.* [1988] provided further refinements of the conceptualisation, as well as measurement instrument. They made a more clear distinction between the two constructs arguing that perceived quality is a form of general and durable attitude towards the firm, whereas satisfaction is only related to a particular transaction, evolving in perceptions of service quality over time. The relation to disconfirmation paradigm was still supported, however the notion of expectations was conceptualised in a different way than in the satisfaction literature. They emphasize that to assess service quality consumers' expectations reflect their desires or wants (should), whereas in the satisfaction literature expectations are viewed as consumers' predictions of how the delivery of the service is likely to look like (*would* expectations). In line with the assumption that encounter specific satisfaction is in the long run an antecedent of service quality was study by Bitner [1990] and Bolton and Drew [1991a].

Perceived service quality is thereby "the consumer's judgment about an entity's overall excellence or superiority" [Zeithaml, 1987; Parasuraman *et al.*, 1988] and can be conceptualised as a form of attitude resulted in comparison of expectations and perceptions of the service performance. In this sense it is related to satisfaction. Parasuraman *et al.* [1988] argue that "incidents of [transaction-specific] satisfaction over time result in perceptions of service quality. They demonstrate that different is the notion of expectations in conceptualisations of satisfaction and service quality. In the satisfaction literature, customers' expectations are viewed as their predictions of how the service delivery will look like and what is likely to happen during the exchange. Unlike in the service quality literature, expectations are viewed as desires or wants of customers (*should* vs. *would*).

2) Operationalization

The first attempt for developing an instrument to assess the customer's perceived service quality was the study by Parasuraman *et al.* [1985]. This study based on the focus group interviews revealed that customers tend to evaluate the performance of service companies across 10 general dimensions. However no measurement instrument was provided within this study.

In a subsequent study by Parasuraman *et al.* [1988], five dimensions were considered adequate for most service industries: reliability, responsiveness, assurance, empathy and tangibles of the service. The study also provided the measurement instrument, called SERVQUAL battery, which is perhaps, the most recognized and verified measure of perceived service quality. It has been successfully tested in many studies and can be applied with minor modifications to measure perceived quality of most types of services, for instance, retail store performance quality. This instrument treated difference between scores for expectations and perceived performance as the determinant of overall service quality.

However with this study, a debate started whether to use perceptions-only, perceptions-minus-expectations scale, or weighted scales respectively.

Cronin and Taylor [1992] compared three alternatives to original SERVQUAL scale measures: weighted SERVQUAL, performance-only (SERVPERF), and weighted performance-only measure and found that performance-only scale is the best as it explains most variance in overall quality.

In another article by Taylor [1995], the author compares four instruments (SERVQUAL, SERVPERF, and two other scales) to find one that best conforms to attitude theory (i.e. the measure should incorporate customer's perceived weights for attributes of an object or a issue).

Parasuraman *et al.* [1994] have used the notion of "zone of tolerance" to determine the shortfalls of service performance. Zone of tolerance is limited on one side by adequate level of the service quality that customer is willing to accept as minimum, and by expected level that customer believes a good service provider should and can provide.

3) Antecedents

The conceptualisation of service quality proposed by Parasuraman *et al.* [1988] embraces implicitly its antecedents. These antecedents are reliability, responsiveness, competence, access, courtesy, communication, creditability, security, understanding/knowing the customer, and tangibles [Parasuraman *et al.*, 1988].

Authors	Dependent Variable	Antecedents	Consequences
Zeithaml *et al.*, 1996	*Service quality*		Loyalty; Propensity to switch; Willingness to pay more; External response to problem
Parasuraman et al., 1985		Reliability, Responsiveness, Competence, Access, Courtesy, Communication, Creditability, Security, Understanding/Knowing the customer, Tangibles	
Bloemer and Kasper, 1995			Satisfaction
Zeithaml et al. 2000	*E-service quality*	Access, Ease of navigation, Efficiency, Flexibility, Reliability, Personalization, Security/Privacy, Responsiveness, Assurance/Trust, Site Aesthetics, Price Knowledge	

Table 3.1.2. Some antecedents and consequences of perceived service quality.

4) Consequences

Cronin and Taylor [1992] demonstrate that service quality is an antecedent of overall satisfaction and has less significant effect on purchase intentions than satisfaction.

Interesting insight to the nature of the relation to loyalty was conducted by Zeithaml *et al.* [1996]. The study investigated effects of overall service quality assessments on behavioural intentions with respect to three levels of quality scores: below the adequate level, within the zone of tolerance, and above it. There exists positive relation between service quality and loyalty and willingness to pay more, and negative and less significant association for propensity to switch and external response.

Recently, researchers have come to agreement that perceived service quality results over time in overall satisfaction [e.g., Oliver, 1993; Rust and Oliver, 1994; Cronin and Taylor, 1992; Zeithaml, 2001]. Satisfaction thus mediates the link between quality and behavioural consequences of loyalty.

5) e-Service Quality

Most of the studies that we have identified in the existing literature on online transactions deal, in one sense or another, with the issue of quality. There exist differences in the literature between authors of what e-service quality really is and means [Zeithaml *et al.*, 2002]. We have discovered that some authors have a too technical, and thus too narrow, view on the issue of quality, and they consider more the quality of the web site itself [Liu and Arnett, 2000; Loiacono *et al.*, 2002; Ranganathan and Ganapathy, 2002; Aladwani and Padvia, 2002], rather than the quality of the entire service delivered through the electronic medium, as conceptualised by Grönross *et al.*, [2000].

Cox and Dale [2001] argue that most of the dimensions and items of the famous SERVQUAL framework [Parasuraman *et al.*, 1988] developed for physical service environments, are not relevant to assess quality in virtual environments related to e-commerce. On the other hand, Liljander *et al.* [2002] propose that traditional quality dimensions can be adapted to capture the new electronic media, but also that additional dimensions will be needed [Grönross *et al.*, 2000; Zeithaml *et al.*, 2000].

An important contribution to the field is the proposal of a battery of dimensions, along which web users evaluate websites, called e-SERVQUAL [Zeithaml *et al.*, 2000]. Zeithaml *et al.* [2002] found that suggest that customer characteristics such as age, gender, income, experience and technology readiness could influence customer perceptions and evaluations of service delivery through web sites. More precisely, these customer characteristics are suggested to exert a moderating effect on the relationship between drivers of e-satisfaction and behavioural outcomes [Bansal *et al.*, 2003].

Conceptualisation

With respect to definitions that seem to be focused more on the web site alone rather than on the entire quality of the e-service, a good instance of such an approach is the article by Aladwani and Palvia [2002]. They define perceived web quality as "as users' evaluation of a web site's features meeting users' needs and reflecting overall excellence of the web site." In line with this definition, the results of their study uncovered four factors of perceived web quality: technical adequacy (including items such as security, ease of navigation, personalization, speed of loading), specific content (e.g., details about products, customer support), content quality (e.g., information usefulness, accuracy), and appearance (e.g., attractiveness, organization, proper use of fonts, colors, and graphics).

Similarly, the WebQual model of Loiacono *et al.* [2000] focuses on the technical quality of the Web site itself rather than on the quality provided to customers. Loicacono *et al.* [2002] extended and re-assessed the WebQual scale, and found that usefulness, entertainment and response time are primary indicators of website quality and are the most important factors predicting reuse of a website, whereas ease of use and trust are less important.

Probably, the first study that attempts to define e-service quality in a broader sense by Zeithaml *et al.*, [2000] states that e-service quality (e-SQ) is "the extent to which a Web site facilitates efficient and effective shopping, purchasing and delivery" [Zeithaml *et al.*, 2000]. The authors bring evidence that online exchanges do obey their own specific rules. Also, they identify eleven dimensions along which users evaluate service quality being delivered by online retailers. Besides five traditional SERVQUAL dimensions (reliability, tangibles, responsiveness, empathy, and assurance), the authors also suggested ease of navigation, flexibility, efficiency, site aesthetics and price knowledge as specific only for online service quality. Simultaneously, Kaynama and Black [2000] propose seven dimensions derived from the SERVQUAL model: responsiveness, content and purpose, accessibility, navigation, design and presentation, background, personalisation and customisation. Based on these works, Liljander *et al.* [2002] propose five dimensions of e-service quality: user interface (as equivalent of tangibles in SERVQUAL), responsiveness, reliability, customisation and personalisation (equivalent of empathy), and trust (assurance). These dimensions were next used in Van Riel *et al.* [2003].

In their next study, Zeithaml *et al.* [2002] claim to know that e-SQ is multifaceted and includes dimensions such as ease of use, privacy/security, fulfilment/reliability, graphic style, and information availability and content.

Bauer and Hammerschmidt [2002] developed and validated a quality assessment scale for web portals in particular. From their study it can be concluded that Internet users perceive three generic services delivered through a web portal. These services serve as the key dimensions for evaluating portal quality. The first dimension, i.e., security/trustworthiness and basic services, represent as the portal's "hardware" the basic demands of portal users in the sense of minimal conditions. Attractive cross-buying services and added values make up a second dimension representing the "software" (additional services) around the core products. A third dimension used for quality assessment consists of transaction support and relationship building services that have to be facilitated through personalized offers and contents and interactive decision tools.

Based on SERVQUAL and TAM (Technology Acceptance Model) models, Song and Zinkhan [2003] propose that perceived web site quality consists of seven dimensions: interactivity, usability, reliability, content quality, entertainment, privacy and security, and merchant brand image.

Some studies focused on attributes of websites and their relation to loyalty. For instance, Chen et al. [2003] advances eight e-store attributes: relative price, merchandise quality, e-retailer's reputation, customer service, safety, order fulfilment, information quality, and website navigation.

In summary, apart from the aforementioned criteria along which customer evaluate e-service quality, we should mention also some more specific items, such as: ease of navigation [Zeithaml *et al.*, 2000; Gommans *et al.*, 2001], ease of use, graphic style, including items such as layout, colors, and graphics [Ariely, 2000; Hoffman and Novak, 1996; Hoque and Lohse, 1999; Lynch and Ariely, 2000; Montoya-Weiss *et al.*, 2000; Novak *et al.*, 2000; Schlosser and Kanfer, 1999], privacy and security [Hoffman and Novak, 1996], information availability and content [Swaminathan *et al.*, 1999; Zellweger, 1997].

Antecedents

Some authors suggest that website quality is more important for vendors selling high-touch rather than low-touch goods [Lynch *et al.*, 2001].

Zeithaml *et al.*, [2000] found that website performance with respect to responsiveness, personalization and the amount of information and graphics is not linearly related to overall perceived service quality, but there exists optimal level of delivery for these dimensions, below and above which service quality decreases. Moreover, these levels vary among customers.

According to Song and Zinkhan [2003], certain features, such as interface design, information access, and fulfilment policy, of the website itself influence the customer's perceptions of Web site quality.

Van Riel *et al.* [2001] investigated the way consumers evaluated an Internet portal site. They have identified the major components of the online service offer

In a recent case study by Van Riel *et al.* [2003], it has been investigated how customer expectation levels, desired and adequate, towards e-SQ influence perceived performance of e-service. For example, the study brought empirical evidence that significant differences in the levels of acceptable service quality exist between customers with a positive and customers with a negative disposition towards e-services. Furthermore, customers with a more favourable attitude towards e-services are actually less tolerant of poor service quality [Van Riel *et al.*, 2003].

Consequences

Web site quality will positively affect consumers' purchase, repeat purchase, and loyalty [Song and Zinkhan, 2003].

Likewise, other authors also agree that e-service quality will influence purchase intentions, and other forms of loyalty [Bansal *et al.*, 2003], as well as satisfaction [Zeithaml *et al.*, 2002; Wolfinbarger and Gilly, 2002].

Some specific dimensions, or attributes of websites will also influence other constructs. Usefulness, entertainment and response time were the most important factors predicting reuse, whereas ease of use and trust are less important [Loiacono *et al.*, 2002].

3.1.4. Image

Evidence can be found that image is related to customer patronage [Korgaonkar *et al.*, 1985; Granbois, 1981] as well as loyalty in general [e.g. Mazursky and Jacoby, 1986; Osman, 1993] and as such is believed to be "a critical aspect of a company's ability to maintain its market position" [Bloemer *et al.*, 1998]. A favourable store image is thought to lead to store loyalty. The exact relationship between image and loyalty is, however, neither simple nor straightforward and remains a matter of debate.

1) Conceptualisation
Many definitions of image can be found in the marketing literature [Doyle and Fenwick, 1974; Kunkel and Berry, 1968], but it has been most frequently conceptualised as "gestalt", depicting customer's overall impression.

Mazursky and Jacoby [1986] identified three general factors of the store image: merchandise-related aspects, service-related aspects, and pleasantness of shopping at a store.

2) Operationalization
It is trivial to say that image conceptualised as a global impression is a very difficult psychological construct to operationalise.

Basically, there are two streams of operationalization of image. Attribute-based stream adopts that image can be operationalized with a list of attributes [e.g., Osman, 1993]. However, this operationalization is not in line with the real notion of image as a global impression. Image is not just a sum of attributes, but "something more than the mere sum" [Keaveney and Hunt, 1992]. Therefore this operationalization is regarded as failing to capture the richness of image as conceptualised.

Keaveney and Hunt [1992] developed category-based operationalization of image. They have argued that image is formed along the category-based information processing theory. This theory posits that customers will attempt to match a stimulus (e.g., when they start a new session on the website) to a known category stored in memory. The stimulus will be then classified into a class of similar concepts, and based upon the characteristics of the class, the customer will form a judgment about the website. This approach is thus based on experience and prior knowledge related to the domain (e.g., of financial websites).

Bloemer *et al.* [1998] operationalise bank image as the battery of six dimensions: customer contacts, advice, relationship-driven items, position in the market, society-driven items and prices.

3) Antecedents

Bloemer et al. [1998] found that position in the market, prices, advice, customer contacts, as well as society- and relationship-driven factors influence image.

Authors	Dependent Variable	Antecedents	Consequences
Darden and Schwinghammer, 1985; Render and O'Connor, 1976;	Image		Perceived quality
Bloemer et al. 1998		Customer contacts; Advice; Relationship-driven factors; Position in the market; Society-driven factors; Prices	Perceived service quality; Satisfaction; Bank loyalty
Sirgy and Samli, 1989			Loyalty

Table 3.1.3. Some antecedents and consequences of image.

4) Consequences

It is agreed upon in the CS&L literature that image is an antecedent of perceived quality [Darden and Schwinghammer, 1985; Render and O'Connor, 1976].

Sirgy and Samli [1989] show that image has direct positive impact on loyalty. It has been evidenced, however, that this link is mediated by evaluative judgments such as quality perceptions and satisfaction. In line with this assertion is a recent study by Bloemer *et al.* [1998], which reports that image results in perceived service quality, satisfaction and bank loyalty, but its indirect effect on commitment and future intentions through perceived service quality is much more significant than the direct one. In the latter study only "position in the market" had direct positive effect on loyalty.

We haven't encountered any studies that address image in the online loyalty context.

3.1.5. Involvement

Involvement reflects the inherent interest a consumer has in the service/product. Peter and Olson [1996] define it as "consumers' perceptions of importance or personal relevance for an object, event, or activity." Different variations of involvement construct may be found in the literature including, e.g. ego, purchase, product, brand, enduring, situation, response, low, and high involvement. The notion of involvement is conceptually similar to commitment, however most of researchers make distinction between them suggesting that commitment refers to a particular position on a brand in the product class, whereas involvement refers to a general level of interest or concern in an issue without reference to a specific position [Freedman, 1964; Zaltman and Wallendorf, 1983].

Involvement addresses forms of arousal or drive activation, and thus can be considered motivating variable [Beatty *et al.*, 1988].

1) Conceptualisation

The most general term relating to involvement is product involvement. Day [1970] shortly conceptualises it as "the general level of interest in the object or the centrality of the object to the person's ego structure".

Beatty *et al.* [1988] distinguish between two major types of involvement in their model: ego involvement and purchase involvement. Ego involvement is defined as "the importance of the product to the individual and to the individual's self-concept, values, and ego".

On the other hand, purchase involvement relates to "the level of concern for, or interest in, the purchase process triggered by the need to consider a particular purchase." Individual characterized as ego-involved will thus, by definition, feel that a particular product category is highly relevant and closely related to his or her ego, values and self-concept, while purchase-involved customer will care more about the choice decision concerning the purchase of a particular product within that category.

Conceptually similar to ego involvement is enduring involvement. Richins and Bloch [1986] suggest that enduring involvement "is independent of purchase situations and is motivated by the degree to which the product relates to the self and/or the hedonic pleasure received from the product". It is a function of previous experience with the product and the strength of values to which the product is relevant.

Related to purchase involvement is situational involvement, that Bloch and Richins [1983] define as "the degree of involvement evoked by a particular situation such as a purchase occasion". Situational involvement can be influenced by product attributes (e.g., product complexity, product cost, similarity among choice alternatives) as well as situational variables.

Another variety of involvement is response involvement, which represents "the complexity or extensiveness of consumer decision making and thus refers to the consequences of the inner state of being involved" [Bloch and Richins, 1983].

Involvement is an attitude that is viewed as not to be directly related to loyalty. Its significance for determination of loyalty is however unquestionable due to functioning as a moderator.

2) Operationalization

In order to measure ego involvement towards products the following statements can be for instance used on Likert scale with anchors "agree-disagree" or "likely-unlikely": "The brands or types of ... I use say a lot about me", "I can make many connections or associations between my use of ... and experiences in my life". Purchase involvement can be assessed by items: "I am very concerned about what brands of ... I purchase."

The following sample items with 5-point Likert scale were used as measures for product involvement: "Choosing a blank audio cassette is not an important decision for me. (reverse scored)" [Mittal and Lee, 1988; Bloemer and Kasper, 1995], "A blank audio cassette is personally relevant for me" [Verplanken, 1991; Bloemer and Kasper 1995].

3) Antecedents

Authors	Dependent Variable	Antecedents	Consequences
Laurent and Kapferer, 1985	*Involvement*	Perceived product importance and importance of the consequences of a mispurchase; Subjective probability of a mispurchase; The hedonic value of the product; The symbolic value of the product class	
Smith and Rutigliano, 2003		Trust; Confidence	
Dick and Basu, 1994			Loyalty
Beatty et al., 1988	*Ego involvement*		Purchase involvement
Beatty et al., 1988	*Purchase Involvement*		Brand commitment

Table 3.1.4. Some antecedents and consequences of involvement.

Laurent and Kapferer [1985] derived four antecedents for involvement: perceived importance of the product and importance of the consequences of a mispurchase, the subjective probability of a mispurchase, the hedonic value of the product class, and the symbolic or sign value of the product class.

One of the building blocks of engagement is trust and confidence [Smith and Rutigliano, 2003]. Engagement is conceptually close to the concept of product involvement.

4) Consequences

The higher the involvement, the more intense is consumers' post-purchase evaluation of performance, which is a composite of satisfaction [Patterson, 1993]. Dick and Basu [1994] posited that "the higher the involvement in a consumption category, the greater the likelihood of loyalty towards specific offerings within this category."

Beatty *et al.* [1988] have found that ego involvement affects purchase involvement, which in turn is directly related to brand commitment (there is no direct relation between ego involvement and brand commitment).

Bloemer and Kasper [1995] use brand choice involvement along with deliberation scale to determine status of customer satisfaction with respect to latent and manifest satisfaction.

In high involvement services (e.g., restaurant, travel agency), link between satisfaction and loyalty is stronger when customers experience positive emotions, whereas in low involvement settings (e.g., local public services) this moderating effect is not so significant [Bloemer and de Ruyter, 1995].

3.1.6. Trust

Customer's trust is also viewed as an essential ingredient for successful relationships [e.g., Berry, 1995; Dwyer *et al.*, 1987]. Morgan and Hunt [1994] consider trust as the key mediating variable in the exchange relationships. Some researchers have tended to emphasize trust as a confidence in a salesperson rather than towards an organization as a whole [e.g. Crosby *et al.*, 1990]. With respect to online relationships, Hoffman *et al.* [1999] argue that earning consumer trust is "the most effective way for commercial Web providers to develop profitable exchange relationships with online customers".

1) Conceptualisation

Trust has been defined as "customer's confidence in the quality and reliability of the services offered by an organization" [Garbarino and Johnson, 1999] or, similarly, as "the belief that a party's word or promise is reliable and a party will fulfill his/her obligations in an exchange relationship" [Dwyer *et al.*, 1987]. Moorman *et al.* [1993] define it as "a willingness to rely on an exchange partner in whom one has confidence." Morgan and Hunt [1994] capture trust as the perception of "confidence in the exchange partner's reliability and integrity."

In general, two competing approaches, "expectancy" and "behavioural", for the conceptualisation of trust co-exist [Singh and Sirdeshmukh, 2000]. "Expectancy" paradigm relates to individual's expectations about the intentions and/or behaviours of the exchange party and focuses on one's beliefs that the exchange partner will act in a responsible, injurious and showing integrity manner. The latter, "behavioural" approach emphasizes one's intentions to rely on the exchange partner accepting the contextual vulnerability.

Rousseau *et al.* [1998], based on different conceptual definitions found in the sociology, psychology and economics literature, proposed the following consensus definition of trust: "trust is a psychological state comprising the intention to accept vulnerability based on positive expectations of the intentions or behaviours of another."

Based on the aforementioned definition of Rousseau *et al.* [1998], Singh and Sirdeshmukh [2000] propose a multidimensional conceptualisation of trust. They argue that three factors must be taken into account that determine customer's trust. First, determinants such as level of customer's perceived uncertainty about the service performance, its consequentiality for the values customer derives from the service, and the extent to which customer participates in the exchange form situational and contextual factors of the exchange [Sitkin and Roth, 1993]. Next, customer's expectations about competence and benevolence determine overall trust expectations. A third aspect worth observation is that high levels of trust and distrust are extreme points on the trust-distrust continuum. However, their conceptualisation remains to be empirically validated.

Geyskens *et al.* [1996] support the notion deriving from the social psychology literature that trust encompasses two essential elements: trust in the partner's honesty and trust in the partner's benevolence. The former element reflects one's belief that the partner will keep its word, fulfil promised obligations and is sincere, whereas the second component pertains to one's belief that one's partner is interested in one's welfare and will not take unexpected actions that could harm one.

2) Operationalization

Trust scale was developed, for instance, by Sullivan *et al.* [1981]. Furthermore, measures developed by Kumar *et al.* [1995] can be used to determine different types of trust with regard to trust in the partner's honesty and trust in the partner's benevolence. Sample item from their scale for honesty (5 items) is "Our organization can count on the supplier to be sincere", and for benevolence (5 items), "When making important decisions, the supplier is concerned about our welfare".

3) Antecedents

Morgan and Hunt [1994] verified that one's partner opportunistic behaviour as perceived by one is negatively related to trust towards that partner. They also found that meaningful and timely communication between partners and shared values, as the extent to which partners have common beliefs about what behaviors, goals, and policies are important or unimportant, appropriate or inappropriate, and right or wrong, engender trust in one's partner.

In channel member behaviour, trust has been found a mediating variable between power structure and use and satisfaction [Duarte and Davies, 2000].

4) Consequences

For high relational customers the greatest driving force of future intentions are trust and commitment. For low relational (transactional) customers it is satisfaction that mediates between satisfaction component attributes and future intentions; in this case satisfaction is related to trust and commitment but they do not affect future intentions [Garbarino and Johnson, 1999].

Trust has a stronger effect on affective commitment than on calculative commitment [Geyskens *et al.*, 1996]

A direct consequence of trust is believed to be customer commitment [Morgan and Hunt, 1994].

Trust is one of the building blocks of customer engagement [Smith and Rutigliano, 2003].

Authors	Dependent Variable	Antecedents	Consequences
Geyskens *et al.*, 1996	Trust		Affective commitment, Calculative commitment (-)
Morgan and Hunt, 1994		Shared values, Communication, Opportunistic behavior (-)	Commitment, Cooperation, Functional conflict, Uncertainty (-)
Gruen *et al.*, 2000			Participation, Coproduction
Anderson and Weitz, 1989			Commitment
Doney and Cannon, 1997		Supplier size, Supplier's willingness to customize, Salesperson expertise,	Future intentions
Garbarino and Johnson, 1999		Encounter-specific satisfaction	Future intentions

Table 3.1.5. Antecedents and consequences of trust.

5) e-Trust

Trust, security, privacy, assurance are conceptually similar concepts that play an important role in many e-loyalty studies that we have consulted [e.g., Zeithaml *et al.*, 2002].

Hoffman *et al.* [1999] posit that trust towards online e-commerce sites arises from customers' perceived lack of control over the access others have to their personal information during the online navigation process. They differentiate between two dimensions of online trust: environmental control, and secondary use of information control. Environmental control affects consumers' perceptions of the security of online shopping [e.g., threat of taking over of credit card number by a hacker], while secondary use of information reflects the consumer's perceived ability to control the use of personal information for other purposes, subsequent to the transaction during which the data was originally collected. Hoffman *et al.* [1999] posit that trust to e-commerce sites is best achieved by allowing the balance of power to shift toward a more cooperative interaction between the online business and its customers. Their conceptualisation of trust is thus rather narrowed technologically oriented.

Trust towards websites has been found to matter more than site quality and experience of positive emotions in explaining customer e-loyalty and purchase intentions [Lynch *et al.*, 2001]. Similarly, in an analysis of various website characteristics, Ranganathan and Ganapathy [2002] it has been found that security and privacy are better predictors of online purchase intent than features such as information content, and design. In contrast, Liu and Arnett [2000] argue that security is only a necessary condition of web site design it cannot attract customers and promote electronic marketing activities.

3.1.7. Commitment

The concept of commitment reveals numerous times in the relationship marketing literature and is commonly agreed to be one of the essential ingredients for successful, long-term relationships [e.g., Dwyer *et al.*, 1987]. Morgan and Hunt [1994] strongly maintain that commitment and trust are the key mediating variables, and when both of these constructs exist in a relationship, they will directly lead to cooperative behaviours that will be conducive to relationship marketing success. Its definitions do not vary significantly across different scholars, however there still exists no consistency or agreement towards its modularisation as well as, consequently, its operationalization within the context of customer loyalty studies. Altogether, its notion as a separate construct has been proven since Staw's [1977] criticism, for instance by Gundlach *et al.* [1995].

Before the notion of commitment got settled in the field of marketing research, it had been extensively studied in the social exchange, marriage and organizational commitment literatures. Originally it was adopted by marketing scientists from a number of sources, but the major developments and recently accepted conceptualisations of this construct stem from the organizational commitment literature. Jacoby and Chestnut [1978] recognized commitment as providing "an essential basis for distinguishing between brand loyalty and other forms of repeat purchase behaviour". Definitions exist that tend to perceive commitment as close to involvement, e.g. Kelley and Davies [1994] define it as "an individual's identification with and involvement in an organization".

1) Conceptualisation

Commitment reflects in general a channel member's intention to continue the relationship and has been typically defined as "an exchange partner believing that an ongoing relationship with another is so important as to warrant maximum efforts at maintaining it" [Morgan and Hunt, 1994] or "a buyer's enduring desire to continue a relationship with a seller accompanied by his willingness to make efforts to maintain it" [Bloemer *et al.*, 2000], "an enduring desire to maintain a valued relationship" [Moorman *et al.*, 1992] or "an implicit or explicit pledge of relational continuity between exchange partners" [Garbarino and Johnson, 1999]. These last authors use the same definition as Moorman *et al.* [1992] but they distinguish between four components of commitment.

There exists clear evidence that commitment is an essential ingredient for successful long-term relationships. It is considered as an ultimate attitudinal aspect of loyalty.

An effort to structuralize the notion of commitment as a multidimensional construct was taken by several authors. In particular, a major contribution to the field was the study by Allen and Meyer [1990], in which the authors proposed three-component model of organizational commitment consisting of continuance, affective and normative components. Those components are believed to arise from different motivations that employees have towards maintaining the relationship with their organization. According to this model, affective commitment is defined as a positive emotional attachment, as the extent to which the individual is bonded with the organization on the basis of how favourable it feels about the organization. Continuance, also called calculative, commitment is based on the self-interest stake or a pledge in a relationship, as the degree to which the employee is psychologically bonded to the organization on the basis of perceived costs (e.g. financial, social) associated with leaving the organization. Normative commitment stems from a person's sense of moral obligation toward an organization, and depicts the degree to which the individual is bonded to the organization on the basis of the perceived moral obligations to maintain the relationship [Allen and Meyer, 1990; Gruen *et al.*, 2000]. The three motivations for staying in a relationship can be thus summarized in the statement that affectively committed employees remain because they want to, those with strong continuance commitment because they need to, and those with strong normative commitment because they feel they ought to do so. Each component develops independently of the others as a function of different antecedents. It is also possible that employees can experience each of these states to varying degrees. Nevertheless, the organization researchers agree that a strong link between commitment and turnover exists, and employees who are strongly committed are those who are least likely to leave the organization.

Based on the prior research of Allen and Meyer [1990, 1991], Gundlach *et al.* [1995] adopted a similar three-component model of commitment between manufacturers and distributors in the behavioural simulation study. The instrumental, also called input, component reflects "an affirmative action taken by one party that creates a self-interest stake in the relationship and demonstrates something, more than a mere promise". It can be thus viewed as some form of investment, pledge or a dedicated allocation of resources, which are specific to the particular relationship and in this way prevent it from fading away. The second component pertains to commitment as an attitude and corresponds to Allen and Meyer [1991] affective component. Within this dimension, commitment may be described as psychological or affective identification and affiliation, apart from its instrumental view. In order to make a clear distinction between attitudinal component and other constructs, such as involvement, motivation, identification or loyalty the authors focused on behavioural intentions meaning of commitment. Thereby, it can be "operationalized in terms of future resources commitment and investments" (*ibid.*). A committed party exhibits then long-term investment intentions and provides the foundation for developing confidence in the stability of the relationship. With respects to the third, temporal dimension, commitment's core elements are its durability and consistency over time, which result in long-term commitment.

They next identified two input dimensions: creditability, i.e. the magnitude of resources pledged by both parties and proportionality of these resources being the second dimension.

In spite of the huge body of evidence on the relevance and validity of the multi-dimensional model of organizational commitment [e.g., Allen and Meyer, 1996], relatively few researchers to date, if any, adopted this approach and verified its scale in the context of buyer-seller relationships, instead conceptualising commitment as the unidimensional construct. However, as Geyskens *et al.* [1996] argue, use of global commitment scale, that does not capture different motivations to continue a relationship, could confound or mask different, and possibly even opposite effects on affective vs. calculative commitment. They emphasize thereby the need of distinguishing between those two facets of commitment in subsequent studies and practice.

2) Operationalization

Measurement instruments reflect nuances in conceptualisation and thus are often designed on ad hoc basis. They are rather context-specific.

The measurement of unidimensional model of commitment can be carried out with the scale used for instance by Bloemer and Kasper [1995].

The measurement instrument for organizational commitment as conceptualised by Allen and Meyer [1990] was developed by Allen and Meyer [1991] and next applied with minor refinements in many subsequent studies [e.g., Iverson and Buttigieg, 1999; Gruen *et al.*, 2000; Geyskens *et al.*, 1996] showing good construct reliability and validity. This instrument operationalises affective commitment with 5-8 (depending on the study) item Likert scales, with sample items "I feel a strong sense of belonging to my organization", or "This organization has a great deal of personal meaning for me". Scale for continuance commitment includes five to eight items, for instance, like "Too much in my career would be disrupted if I decided I wanted to leave my organization", "I feel that I have too few options to consider leaving this organization". Finally, normative commitment can be assessed with several items like "I do not believe that a person must always be loyal to his or her organization", and "One of the major reasons I continue to work for this organization is that I believe that loyalty is important and therefore feel a sense of moral obligation to remain". All scales have anchors "strongly disagree – strongly agree".

Similar to the aforementioned instrument for measuring affective and calculative commitment was developed by Kumar *et al.* [1994].

3) Antecedents

It is viewed as a direct consequence of service quality and satisfaction [Kelley and Davies, 1994], and direct antecedent of customer service recovery expectations, word of mouth, price sensitivity and repeat purchasing [Bloemer *et al.*, 2000].

An interesting insight into the study of commitment was provided by Geyskens *et al.* [1996]. They found that trust was positively related to affective commitment, and negatively linked to calculative commitment.

Authors	Dependent Variable	Antecedents	Consequences
Allen and Meyer 1996	Affective commitment	Job satisfaction, Job involvement	Turnover intentions
Gruen et al. 2000		Recognition, Dissemination of knowledge	Participation, Coproduction
Iverson and Buttigieg 1999		Job expectations, Job values, Positive affectivity, Work motivation	Turnover intentions, Absenteeism, Organizational change
Iverson and Buttigieg 1999	Continuance commitment	Job security, Job opportunities,	Turnover intentions, Organizational change
Iverson and Buttigieg 1999	Normative commitment	Kinship responsibilities, Job hazards, Routinization	Turnover intentions, Absenteeism
Kelley and Davies 1994	Commitment	Service quality, Satisfaction	
Morgan and Hunt 1994		Trust, Shared values, Relationship benefits, Termination costs	Cooperation, Acquiescence, Propensity to leave
Swan and Oliver 1991		Trust, Shared values, Termination costs, Relationship benefits	
Gundlach et al. 1995		Creditability, Proportionality	Long-term commitment
Macintosh and Lockshin 1997		Salesperson trust	Store attitude, Purchase intentions
Bloemer et al. 2000		Relationship proneness, Social affiliation	Word-of-mouth, Price sensitivity, Repeat purchasing

Table 3.1.6. Some antecedents and consequences of commitment.

Morgan and Hunt [1994] prove that termination costs and relationship benefits, shared values as well as trust are direct antecedents of commitment.

4) Consequences

Morgan and Hunt [1994] argue that commitment and trust engender cooperation.

Commitment is widely believed to be a psychological state that ultimately determines behavioural outcomes of true customer loyalty, as repeat purchasing, word-of-mouth communication, and complaint behaviour. Its ability to explain variance in future intentions is also the most significant one of all the other customer attitudes [Garbarino and Johnson, 1999].

3.1.8. Outcomes

Each of the aforementioned constructs is critical for successful long-term relationships and may be viewed as a necessary precondition for true, enduring loyalty. However, in order that true loyalty to exist, these constructs must be accompanied by behavioural outcomes.

Dick and Basu [1994] propose in their framework that search motivation, word-of-mouth communication, and resistance to counterpersuasion are also important consequences of loyalty.

Most authors would agree that the most important outcomes of successful customer-seller relationships are word-of-mouth communication, purchase intentions, price sensitivity, repeat patronage, and complaint behaviour [Bloemer *et al.*, 2000]. These outcomes make up so-called behavioural-intentions battery [Zeithaml *et al.*, 1996]. Some researchers emphasize value of search motivation.

Willingness to recommend, or word-of-mouth communication behaviour, reflects "post-purchase communication by consumers" [Oliver, 1980]. Other definitions treat it as "volitional information dissemination."

Repeat purchasing has been already discussed as the most tangible effect of customer loyalty [Jacoby and Chestnut, 1978].

As regards complaint behaviour, we can discriminate between internal and external complaint behaviour.

The motivation to search for information about alternative brands/service providers may be seen as a function of consumer's perceived benefits and costs of search activity. Costs of the activity can be associated with lost time or money, as well as psychological inconveniences due to delayed gratification. Perceived benefits, in turn, are viewed to be reduced in case when consumers have high relative attitude and/or are engaged in repeat patronage [Ratchford, 1980; Dick and Basu, 1994].

Dick and Basu [1994] cite evidence from a number of studies that the search motivation decreases as perceived satisfaction, repeat purchase, experience and learning increase. For instance, Furse *et al.* [1984] report that buyers of new cars, who are low search motivated, have the most purchase experience and are more satisfied than high search consumers, who are least satisfied with previous purchases and have the lowest confidence in their ability to choose.

3.1.9. Moderating variables

Moderators, or in other words, moderating variables, are phenomena or concepts that extend attitudes-behaviour model. Moderating variables have an effect on the strength of the relationship between attitudes and behavioural-intentions battery, but do not comprise factors that are directly related to loyalty towards a particular service provider. However their incorporating provides more descriptive as well as predictive power into the model.

Sekaran [2003] defines a moderating variable a one that has "a strong contingent effect on the independent variable-dependent variable relationship." The presence of the moderating variable modifies thus the original relationship between the independent and the dependent variables. There are lots of factors that are regarded as moderators; here we give only some examples thereof.

Some of the most essential moderating variables are situational and contextual factors. For instance, Dick and Basu [1994) argue that relationship between attitudes and repeat purchasing is moderated by social norms and situational factors.

Many authors agree that situational factors, like store location, affect loyalty, but they are in general considered to drive spurious loyalty. Some authors indicate age of customer-company relationship as a moderating variable.

Sociodemographics and personal characteristics seem also to moderate the relationships between many of the aforementioned constructs. For example, the recent work by Bloemer et al. [2001] demonstrate that relationship proneness is a very important factor influencing consumers' attitudes and behavioural store loyalty. They suggest incorporating of other personality-related constructs as they appear to be important antecedents of loyalty.

Furthermore, Bloemer and De Ruyter [1999] indicate that, especially in the case of high involvement services, positive emotions like excitement, enthusiasm, inspiration experienced during the service delivery strengthen the link between satisfaction and loyalty. Similar findings are pointed out by Westbrook [1987], who stated that consumers are more likely to engage in word-of-mouth communication when they experience notable emotional experiences.

The relationship between satisfaction and service loyalty is also moderated by value attainment and experiencing positive mood [de Ruyter and Bloemer, 1998]. Gundlach *et al.* [1995] found that creditability of resources pledged to a relationship is conducive to development of relational social norms, such as solidarity, mutuality, flexibility which, in turn, reinforce commitment itself and are related positively to future commitment intentions.

With regard to online customer behaviour, Ranaweera *et al.* [2003] propose that customer characteristics, such as technology readiness, flow, purchase involvement, demographics, trust disposition, and risk perception moderate the link between website satisfaction and behavioural and intentional outcomes.

With regard to sociodemographics, Simon [2001] explored the impact of culture and gender on perception of websites and found that perception and satisfaction differences exist between the cultural clusters - such as Asia, Europe, Latin & South America, and North America - and gender groups within those cultures. To be more detail, the perceptions of the Europeans and North Americans were found to be similar, as were the perceptions of the Asian and Latin/South Americans. Furthermore, his qualitative analysis indicates that females within certain cultures have widely different preferences from males regarding web site attributes. In particular, the study discovered that female responses for perception were significantly lower than male responses on all four websites that the author considers, and as a result, male would be more satisfied with web sites than females. The author argues that this finding is consistent with previous studies, which indicate that female use a more comprehensive information-processing scheme and are less fascinated with technology than males.

A number of studies suggest that demographics can act as moderators between e-satisfaction and customer e-loyalty [e.g., Anderson and Srinivasan, 2003]. The sociodemographics can also play an important role in the formation of the feelings and opinions towards websites [Ranaweera *et al.*, 2004]. Important website features could in this respect include the design of the website, look and feel of the pages, or style and amount of graphics. It can be expected, as an example, that younger web surfers could favour fancy animations or music add-ons more than older users who would be irritated with the page overload. The technology in which the website is created and presented, being either simple and familiar HTML or trendy Flash technology can also affect the perception of navigation patterns. These considerations are supported by Simon [2001], who found that as much as 84% females preferred sites that are less cluttered, with minimal use of graphics and sites which avoid multiple levels of sub-pages to drill through, whereas 77% males, on the other hand, indicate that sites making extensive use of graphics and animated objects are clearly their preference.

The relationship between e-satisfaction and e-loyalty has been found to be moderated by consumers' individual level factors, and by firms' business level factors [Anderson and Srinivasan, 2003]. As we have already mentioned, factors that accentuate the impact of e-satisfaction on e-loyalty are convenience motivation and purchase size, whereas inertia suppresses this impact. On the business' level factors, both trust and perceived value significantly emphasize the impact of e-satisfaction on e-loyalty [Anderson and Srinivasan, 2003].

3.2. Causal modelling approaches in the CS&L research

In this section we give a brief overview of most prevalent research techniques and methodologies applied today in the Customer Satisfaction and Loyalty modelling. The purpose is to present mainly the most relevant characteristics, applicability and limitations of these techniques.

The application area of modelling is usually referred to as *causal modelling* [Bagozzi, 1982], as the notion of causality, more or less explicitly, is involved in every study, and it is implied in such studies that the direction of causality has been established [Hulland *et al.*, 1996].

In general, in the literature on marketing modelling sometimes a distinction is made between first and second generation of multivariate analysis techniques. With regard to first generation statistical techniques, it is a general term relating to correlation based analyses methods like linear regression, logit, ANOVA, MANOVA, etc. As regards second generation techniques, Fornell [1982] refers with this name to causal modelling and suggest a number of ways in which these techniques are superior to first generation techniques, including: (i) the explicit inclusion of measurement error, (ii) an ability to incorporate abstract and latent constructs, (iii) an opportunity to not only combine theory and data, but also to confront theory with data. Clearly, first generation techniques preceded chronologically techniques labelled as second generation techniques, which can be seen as more involved and advanced data analyses approaches.

In order to make it clear we will consider three techniques now in more detail, namely (i) linear regression, (ii) LISREL models, and (iii) Partial Least

Squares models. These three techniques can be considered as standard approaches applied in the CS&L research.

3.2.1. Regression models

We begin with the regression models as they are, mathematically viewed, the most straightforward, first generation techniques. In CS&L studies, linear regression models are used for instance in [Naumann and Giel, 1995; Oliver, 1996; Rundle-Thiele and Lockshin, 2001].

1) Characteristics
Regression models are regarded as an instance of first generation techniques [Gefen *et al.*, 2000]. These techniques require researchers to analyse the item loadings on the latent variables separately from the linkage of the independent variables to the dependent variable [Gefen *et al.*, 2000].

2) Specification
Prior to any analysis, a researcher must usually establish some set of hypotheses for a theoretical model at hand. This step is of course common for all techniques considered here. Each hypothesis refers typically to the existence of a causal relationship between two constructs. Sometimes, also the nature of this relationship is hypothesized, in terms of positive or negative effect.

In the first step, tests to ensure the reliability and validity of each construct based on observed variables making up the measurement scale must be performed. The tests most frequently used in this respect refer to construct validity and reliability, which can be established by factor and reliability analyses, respectively [Straub, 1989]. Often Principal Components Analysis is used for this purpose too [Gefen *et al.*, 2000].

3) Estimation
Once the reliability and validity of each scale has been established, an index must be created for each latent variable on the basis of observed variables retained after the tests of validity and reliability. This index is typically obtained by averaging over the observed variables in the scale.

In order to test the hypothesized relationships, in the next step, a series of linear regression analyses, each for one dependent variable "regressed" on other variables, which are hypothesized to be related to this variable, is carried out.

Therefore, in linear regression a series of several unrelated analyses are required: (i) examining how items load on respective constructs via factor analysis, and subsequently, (ii) a separate examination of the hypothesized paths, run independently of the factor loadings.

4) Evaluation

Linear regression is a technique that uses least squares fitting in order to determine the best parameters in a linear function describing a set of data points. As known, least squares fitting method is a mathematical procedure for finding the best fitting curve to a given set of points by minimizing the sum of the squares of the offsets ("the residuals") of the points from the curve.

5) Interpretation
Interpretation is achieved as in the case of simple linear regressions. The linear coefficients represent the change in the regressed variable as a result of the increase in the corresponding repressor variable by one unit.

6) Limitations
The most important limitation of simple regression is that it does not allow for modelling of the measurement model. Furthermore, every regression equation is estimated and evaluated in isolation from other equations, so that it is not possible to test the all relationships in a single statistical test,

Further limitations include: linear relationships, strict restrictions on data distribution, are susceptible to outliers, etc.

Given their limiting power, the use of regression models is less often applied in favour of the Structural Equation Modelling, to be addressed next.

3.2.2. Structural Equation Modelling

In this section we cover undoubtedly the most prevalent approach applied nowadays in the CS&L research, and in "causal" modelling in general, namely the Structural Equation Modelling, or SEM. According to Steenkamp and Baumgartner [2000], SEM has much to offer in the area of theory building. Hulland *et al.* [1996] review the use of causal models published during 1980-1994 in marketing research.

Structural Equation Modelling techniques are second-generation techniques [Fornell, 1982]. On the contrary to first generation techniques, second generation data analysis techniques enable researchers to answer a set of interrelated research questions in a single, systematic, and comprehensive analysis by modelling the relationships among multiple independent and dependent constructs simultaneously [Gerbing and Anderson, 1988]. This capability of holistic and simultaneous analysis is what differs greatly from most first generation regression techniques, which can only analyse one layer of linkages between independent and dependent variables [Gefen *et al.*, 2000].

Structural Equation Modelling involves three primary components [Chin, 2000]:

(i) *Indicators*, also often called manifest variables or observed measures/variables. For questionnaire-based research, each indicator represents a particular question.

(ii) *Latent variables*, or construct, concept, factor. Latent variables are used to represent phenomena that cannot be measured directly.

(iii) *Path relationships* (correlational, one-way paths, or two way paths).

SEM assesses not only the structural model, but in the same analysis, also evaluates the measurement model. This combined analysis enables measurement errors of the observed variables to be analysed as an integral part of the model, and factor analysis to be combined in one operation with the hypotheses testing. In consequence, a researcher achieves a more rigorous analysis of the proposed research model and a better methodological assessment tool [Bollen, 1989].

In fact, SEM modelling is a family of techniques that are founded on different objectives of analyses, statistical assumptions, and the nature of fit statistics they produce. One sub-family of these techniques are covariance-based structural equation models, the most known of which is LISREL (Linear Structural Relations) modelling. Often, the term Structural Equations Modelling is used in the CS&L literature interchangeably to denote LISREL models in particular. Other, less known approach is based on partial least squares principle; hence its name is Partial Least Squares, or PLS, modelling. Both LISREL and PLS have many features in common, nevertheless we will discuss them apart, first LISREL modelling, and later PLS.

3.2.2.1. Covariance-based structural models

LISREL is the name of a popular computer program for covariance-based structural equation modelling developed by Karl Jöreskog. Other names that apply to the same principle as LISREL modelling include methods such as LISCOMP, EQS or EzPath, which are also names of software packages.

1) Characteristics

The objective of LISREL modelling is to minimize the difference between the sample covariances and the covariances implied by the theoretical model.

2) Specification

Table 3.2.1 contains terms and elements in structural model used by convention in the LISREL language.

Symbol	**Name**	**Definition**
		Variables
η	Eta	endogenous latent constructs
ξ	Xi or Ksi	exogenous latent constructs
ζ	Zeta	the error terms in structural equations, normally distributed
		Coefficients
B	Beta	paths connecting one endogenous η to another
Γ	Gamma	paths connecting exogenouos ξ to endogenous η constructs
		Covariance matrices

Φ	Phi	shared correlation matrix among ξ
Ψ	Psi	shared correlation matrix among the error terms of the η

Table 3.2.1 Notation for structural model.

In addition to the structural model, the measurement model consists of X and Y variables, which are observations or the actual data collected. More specifically, X and Y are the measures of the exogenous and endogenous constructs, respectively. Each X should load onto one ξ, and each Y should load onto one η. The notation used for the measurement model is given in Table 3.2.2.

Symbol	Name	Definition
		Variables
$\mathbf{y}$		Observed indicators of η
$\mathbf{x}$		Observed indicators of ξ
ε	epsilon	Measurement errors for y
δ	delta	Measurement errors for x
		Coefficients
Λ_y	Lambda y	the path between an observed variable Y and its η, i.e., the item loading on its latent variable.
Λ_x	Lambda x	the path between an observed variable X and its ξ, i.e., the item loading on its latent variable.
		Covariance matrices
Θ_ε	theta epsilon	the error variance associated with this Y item, i.e., the variance not reflecting its latent variable η.
Θ_δ	theta delta	the error variance associated with this X item, i.e., the variance not reflecting its latent variable ξ.

Table 3.2.2 Notation used for measurement model.

For any proposed theoretical model, the above-mentioned components can be portrayed with a path diagram. Path diagram represents a set of structural equations. Therefore, another practical distinction between first and second generation techniques is the special diagrammatic syntax used in SEM.

Let us now consider an example of a path diagram representing a theory in Figure 3.2.1. According to the convention used in SEM modelling, indicators are in SEM paths usually represented as squares, whereas latent variables are normally drawn as circles. Relationships between latent variables, and between latent and observed variables are defined using arrows.

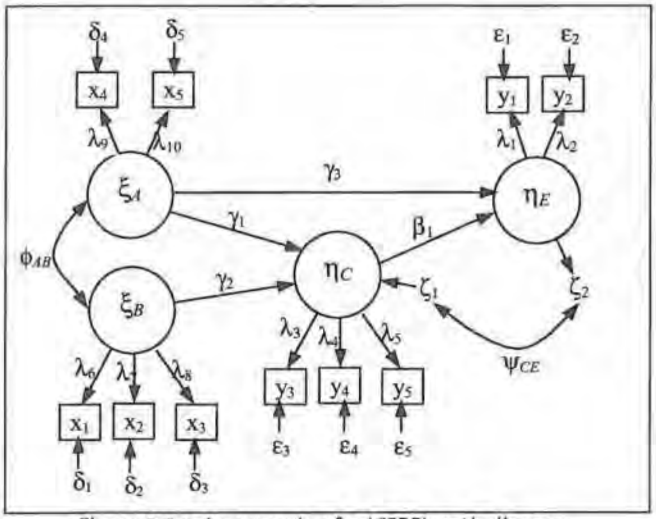


Figure 3.2.1 An example of a LISREL path diagram.

In this example, ξ_A and ξ_B are exogenous, whereas η_C and η_E are endogenous constructs. In line with the notation, the example model in Fig. 3.2.1 is on the structural level equivalent to the following set of two linear equations:

$$\begin{aligned} \eta_C &= \gamma_1\xi_A + \gamma_2\xi_B + \zeta_1, \\ \eta_E &= \gamma_3\xi_A + \beta_1\eta_C + \zeta_2, \end{aligned} \tag{3.1}$$

where γ_1, γ_2, γ_3, and β_1 are linear coefficients for paths connecting the latent constructs, and indicate the relative strength of the relationships. The variables ζ_1 and ζ_2 in these structural equations are the error terms assumed normally distributed. These structural error terms reflect the effects of unmeasured variables, which lie outside the model. The matrix ψ_{CE} is the correlation matrix among the errors terms of the endogenous variables, and ϕ_{AB} is the shared correlation matrix among the exogenous variables. This covariance comes from common predictors of the exogenous constructs which are not accounted for in the model under consideration.

On the level of the measurement, the following two equations denote operationalization, for instance, of the construct ξ_A:

$$\begin{aligned} x_4 &= \lambda_9\xi_A + \delta_4, \\ x_5 &= \lambda_{10}\xi_A + \delta_5, \end{aligned} \tag{3.2}$$

where δ_4 and δ_5 are the measurement errors. The measurement of other latent constructs is modelled in a similar manner. As we can see, the statistical model in LISREL models requires the specification of the linear form of the relationship. So, generally, a LISREL model can be stated as follows:

$$\begin{aligned} \boldsymbol{\eta} &= \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \\ \mathbf{x} &= \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \\ \mathbf{y} &= \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \end{aligned} \tag{3.3}$$

where the notation is the same as in Tables 3.2.1-2.

LISREL modelling makes a lot of assumptions. The fundamental assumption in LISREL models is that the error term in each relationship is uncorrelated with all the independent constructs, i.e., ξ is uncorrelated with ζ. The authors of the program argue that "studies should be planned and designed, and variables should be chosen so that this is the case. Failure to do so will lead to biased and inconsistent estimates (omitted variable bias) of the structural coefficients in the linear equations and thus invalidate the testing of the theory. Omitted variables bias is one of the most difficult specification errors to test" [Jöreskog and Sörbom, 1993]. The errors are assumed to have a multivariate normal distribution.

3) Estimation

The objective of LISREL modelling is to minimize the difference between the sample covariances and those predicted by the theoretical model. This estimation for any given model is iterative. More precisely, during the estimation, such values of correlation coefficients are looked for that can "explain" the "true" correlation matrix as close as possible, and so as to minimize the difference between the sample correlation matrix and the "true" correlation matrix.

A LISREL model may be estimated by seven different methods, i.e., Instrumental Variables (IV), Two-Stage Least Squares (TSLS), Unweighted Least Squares (ULS), Generalized Least Squares (GLS), Maximum Likelihood (ML), Generally Weighted Least Squares (WLS), and Diagonally Weighted Least Squares (DWLS), of which Maximum Likelihood is most popular.

4) Evaluation

The statistical validation is based on the rules of thumb. We must notice there is no mathematical or other means whatsoever to establish the right levels of the goodness-of-fit indices [Gefen *et al.*, 2000; Nunally, 1967].

Furthermore, the goodness of fit, such as χ^2, can test the restrictions implied by the model. In other words, the statistical goal in covariance-based SEM is to show that the operationalization of the theory under examination is corroborated and not disconfirmed by the data [Gefen *et al.*, 2000; Bollen, 1989]. The chi-square test does not confirm a model – it merely fails to reject it.

5) Interpretation

As a consequence, LISREL models should be used as a confirmatory and not as an exploratory method [Bullock *et al.*, 1994; Rigdon, 1998]. As such, it can be applied to show that the theoretical hypotheses developed in a study are supported and plausible given the data.

The explanatory potential of LISREL modelling comes for the most part from the estimation procedure alone, and not from the resulting model itself. This is actually the core of the explanation in LISREL modelling in that the values predict or explain the covariance of observed variables. But when the estimation of these coefficients is finished, and the model is ready to, then the modal provides very little explanatory value. Furthermore, it is also difficult to speak about prediction capabilities in this sense that: 1) it is not possible to infer the values of the latent variables on basis of some particular observations, e.g., taking into account values of the observed variables in one case, and 2) it is not possible to predict possible values of other observed variables [Chin, 1998]. It is also in this sense that Steenkamp [2000] argues "SEM is more focused on explaining marketing phenomena than on predicting specific outcome variables."

6) Limitations

Although now in place for many years, LISREL modelling still poses significant methodological gaps and bottlenecks [Rigdon, 1998]. There are lots of issues that can be by many researchers viewed as severe shortcomings, and by others merely as mild assumptions. We shall address some of them at this point without delving into details. We should note also that some of these limitations are not solved because they cannot be solved due to inherent nature of these methods, while other problems have been tackled for some time now.

One of the most crucial characteristics that in fact strongly limit the practical value of LISREL models are the issues of the linearity and multinormality. Many authors widely agree that these requirements "limit the use of the approach for calibrating response models" [*e.g.*, Lilien *et al.*, 1992, Babakus *et al.*, 1987]. We must admit that as regards the linearity, procedures for incorporating non-linearity have been proposed and developed [see e.g., Kenny & Judd, 1984; Ping, 1996]. Furthermore, there are usually strong restrictions on the distribution of the data. In most typical case, i.e., when the Maximum Likelihood estimation is performed, the data are required to have multivariate normal distribution. The data can have also other distribution than multivariate normal, but then the estimation procedure applied must be different, e.g., generalized least squares [Rigdon, 1998].

The normality of data can be influenced by one or more extreme observations that are quite different from the rest of data, known as *outliers*. These extreme values can seriously impact the results [Shumacher and Lomax, 1996]. Although multivariate outliers are difficult to detect, some methods exist that cope with this problem [*idem*].

Another potential problem relates to the issue that the data cannot be, in principle, missing. When the data are missing, several procedures can be applied, but they have influence on the results. For instance, listwise deletion can have a tremendous effect on sample size, and may induce additional bias. Pairwise deletion, in turn, can lead to input covariance matrices that behave poorly in statistical sense, and is inconsistent with some estimation methods. The missing data problem is currently experiencing rapid advances [Rigdon, 1998], but we note that by its very nature there is no direct way to handle missing data problem.

LISREL modelling is not suitable when dependent variables are categorical, they must be, at least conceptually, continuous. Potential solutions involve introducing dummy variables also in regression.

The constructs have to be measured with at least three indicator variables [Baumgartner and Homburg, 1996, p. 144]. With only one indicator variable for a construct, the measurement error variance may not be identified, and thus its value must be set a priori.

Another important point is that statistical tests are susceptible to sample data sizes. For this purpose, LISREL uses a heuristic - at least 200 cases or 5 to 10 times the number of parameters estimated [Rigdon, 1998]. Gefen *et al.* [2000] argue that the minimum number of cases is 100-150 cases. This requires much larger sample sizes than are needed for regression models.

One of the most serious limitations is the fact that models can often be not identified. A model is identified if it is consistent with one unique set of parameter estimates. Moreover, there is real possibility that the model will fail to produce interpretable results; even when the model is well identified, it is possible that the iterative estimation procedure will fail to converge on a solution, or may converge in a solution that involves unacceptable coefficients [Rigdon, 1998].

LISREL models are also susceptible to *over-fitting*. This problem concerns the situation when researchers modify a rejected model in order to achieve good fit. The new modified model could often have better fit, but at the expense that it will not be replicable in other studies, and so not yield the true model [Steiger, 1990].

Another limitation is that many distinct models, different from each other by the causal paths and managerial implications, can be to the same extent supported by the same data. Such models are called *equivalent models* [Stelzl, 1986].

For many researchers, it is frustrating that the scores for latent constructs are indeterminable [Rigdon, 1998]. That is one of the most important reasons why LISREL cannot be deemed a predictive technique.

Last but not least, in order to use LISREL modelling and knowledgeably interpret the results, one must have a background in matrix algebra, the more so for instance, to understand outcomes when they are negative, or to tell when a seemingly negative result is actually a positive one [Rigdon, 1998]. This puts a heavy burden on marketing researchers and especially on marketing practitioners, who will find the LISREL technique too difficult to adopt it in their research toolkit.

3.2.2.2. Partial Least Squares

Compared to LISREL, the Partial Least Squares, or PLS, is much less known technique and much less often applied in the CS&L research. It has nevertheless been used in a variety of applications, see for instance [Barclay, 1991; Fornell, 1992; Fornell and Cha, 1994; Hulland and Kleinmutz, 1994; Johnson and Fornell, 1987]. Fornell [1992] argued for using PLS as a method for estimating the latent variable Customer Satisfaction Index (CSI) models. More recently however, O'Loughlin and Coenders [2002] compared LISREL modelling versus PLS modelling in the Customer Satisfaction context and found that thanks to recent advances, new procedures in LISREL were advantageous over PLS. According to other authors, the PLS approach can be argued to be more suitable than LISREL depending on the researcher's objectives, properties of the data at hand, or level of theoretical knowledge and measurement development [Chin, 1998].

1) Characteristics

The conceptual core of PLS is an iterative combination of principal components analysis relating measures to constructs, and path analysis allowing the construction of a system of constructs [Thompson *et al.*, 1995]. The objective in PLS is, like in regression, to show high R^2 and significant t-values, thus rejecting the null hypothesis of no effect [Thompson *et al.*, 1995].

One of the most important characteristics is that it is possible to use both formative (cause) as well as reflective (effect) observed or manifest variables as indicators of latent constructs.

2) Specification

PLS models consist of three set of relations: a) the inner model, that specifies the relationships between latent variables, b) the outer model that specifies the relationships between the latent constructs and their associated observed variables, c) the weight relations upon which case values for the latent variables are estimated [Chin, 1998].

The inner model depicts in fact the structural model based on substantive theory

$$\boldsymbol{\eta} = \mathbf{B}_0 + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{3.4}$$

where $\boldsymbol{\eta}$ represents the vector of endogenous latent variables, $\boldsymbol{\xi}$ is a vector of the exogenous latent variables, and $\boldsymbol{\zeta}$ is the vector of residual variables (unexplained variance). This inner model can also be specified in terms of the predictor specification [Wold, 1988]:

$$E(?_j \mid \forall ?_i, ?_h) = \sum_i B_{ji} ?_i + \sum_i ?_{jh} ?_h \,. \tag{3.5}$$

So, it is assumed that each latent variable is a linear function of its predictors and that there is no linear relationships between the predictors and the residual

$$E(?_j \mid \forall ?_i, ?_h) = 0 \text{ and } Cov(?_j, \mid ?_i) = Cov(?_j, \mid ?_h) = 0 \tag{3.6}$$

for the indices i and h ranging over all predictors.

The outer model (also referred to as the measurement model) defines how each block of indicators relates to its latent variable. The manifest variables are partitioned into nonoverlapping blocks. In the case of reflective indicators, the relationships can be defined as

$$\mathbf{x} = \boldsymbol{\Lambda}_\mathbf{x}\boldsymbol{\xi} + \boldsymbol{\varepsilon}_\mathbf{x}, \quad (3.7)$$
$$\mathbf{y} = \boldsymbol{\Lambda}_\mathbf{y}\boldsymbol{\eta} + \boldsymbol{\varepsilon}_\mathbf{y},$$

where x and y are the manifest variables for the exogenous ξ and endogenous ζ variables, respectively. $\boldsymbol{\Lambda}_\mathbf{x}$ and $\boldsymbol{\Lambda}_\mathbf{y}$ are the loadings matrices representing simple regression coefficients connecting the latent variables and their measures. The residuals for the measures $\boldsymbol{\varepsilon}_\mathbf{x}$ and $\boldsymbol{\varepsilon}_\mathbf{y}$ can be interpreted as measurement errors or noise.

Predictor specification is assumed for the outer model in reflective mode as follows

$$E(x \mid ?) = \Lambda_x ?, \quad (3.8)$$
$$E(y \mid ?) = \Lambda_y ?,$$

For those blocks in a formative mode, the relationship is defined as

$$? = \prod\nolimits_? x + \mathrm{d}_? \quad (3.9)$$
$$? = \prod\nolimits_? y + \mathrm{d}_?$$

where ξ, η, x and y are the same as those used in Equation 3.7. Π_x and Π_y are the multiple regression coefficients for the latent variables on its block of indicators, δ_x and δ_y and are the corresponding residuals from the regression. Predictor specification is also in effect as

$$E(? \mid x) = \prod\nolimits_? x \quad (3.10)$$
$$E(? \mid y) = \prod\nolimits_? y \,.$$

Finally, we need to define the weight relations, so that the case value for each latent variable is estimated as

$$\hat{?}_h = \sum\nolimits_{kh} w_{kh} x_{kh} \,, \quad (3.11)$$
$$?_i = \sum\nolimits_{ki} w_{ki} y_{ki}$$

where w_{kh} and w_{ki} are the k weights used to form the latent variables' estimates of ξ_h and η_i.

3) Estimation

The estimation of the parameters representing the measurement and path relationships is accomplished using Ordinary Least Squares (OLS) techniques. The goal is to obtain determinate values of the latent variables, and to minimize variance of all dependent variables, whether observed or latent. The estimation process is iterative. The first stage consists of iterative scheme of simple and/or multiple regressions contingent on the particular model which is performed until a solution converges on a set of weights used for estimating the latent variable scores. Once the estimates for latent constructs are obtained, stages 2 and 3 are simple noniterative applications of Ordinary Least Squares regression for obtaining loadings, path coefficients, and mean scores and location parameters for the latent and observed variables.

4) Evaluation
Evaluation of PLS models applies prediction-oriented measures that are non-parametric. This is the consequence of the fact that PLS makes no assumptions on data distribution, other than predictor specification. To assess predictiveness, PLS uses the R-square for dependent latent variables, the Stone-Geisser test for predictive relevance [Geisser, 1975; Stone, 1974], and Fornell and Lacker's [1981] average variance extracted measure. Resampling procedures such as jacknifing and bootstrapping can be used to examine the stability of estimates [Chin, 1998].

5) Interpretation
The interpretation of results from PLS models must be viewed from the perspective of measures used to evaluate the model. Since the main measure is the R-square measure of variance explained we should view the coefficients as a measure how much variance is explained. Of course, the path coefficients and the loadings for reflective indicators are of interest, too.

PLS do not require strong theory and can be used also in exploratory, or theory-building research settings [Chin, 1998]. It is more suited for predictive applications and theory building, in contrast to LISREL [Gefen *et al.*, 2000].

6) Limitations
PLS models require typically complete data sets; missing data can be filled only by imputation, mean substitution, and listwise deletion [O'Loughlin and Coenders, 2002]. Another limitation is the issue of linearity.

McDonald [1996] criticized PLS modelling for difficulty to describe and extreme difficulty to evaluate. He argues that PLS constitutes "a set of ad hoc algorithms that have generally not been formally analysed, or shown to possess any clear global optimising properties" [McDonald, 1996].

3.2.3. Other techniques

There are a lot of other statistical tools and techniques for categorical data in use by CS&L researchers and practitioners. We do not discuss them here in detail since they are either less known or less appropriate because cannot be deemed causal modelling techniques.

One of the techniques that require more attention is generally known as latent class modelling. These techniques are based on probabilistic formalisms, in which the relationships among variables are described by probabilistic distributions. These models have been adapted to path-like models by Hagenaars [1988], by imposing other restrictions on the form of dependencies between variables, e.g., with loglinear functions, and implemented in LCAG - a program for loglinear (path) models with latent variables that handles variables with missing data [*idem*]. Furthermore, Vermunt [1996] proposed further improvements to loglinear path models, some of which have been implemented in the Latent Gold program [Vermunt and Magidson, 2003].

3.3. Conclusions

3.3.1. Conclusions for the e-loyalty study

In Chapter 4, we discuss the Bayesian network approach by the way of application in the e-loyalty research. We would like to emphasize, that for this purpose we use a secondary data set, in which we found several constructs of value for the our considerations based on the review in this chapter. In this conclusion, we present our motivation for the selection of these variables.

Let us start with our conceptualisation of e-loyalty. For the purpose of the first case study in this work, we will define customer e-loyalty as the behavioural and intentional willingness to be a user of a specific website. Our theoretical definition is thus two-dimensional. The behavioural dimension is reflected with an actual online presence by the concept of stickiness. Stickiness is a metric coined to express how attractive, or "sticky", a website is in terms of time spent on it by its visitors [Bansal *et al.*, 2003; Cutler and Sterne, 2000]. Here we apply this term rather to web users. We define operationally stickiness as the average duration of a visiting session at a website; in our study, it is expressed in seconds. In order to rule out the possibility of spurious behavioural loyalty, we argue that the e-loyalty should also account for the subjective opinion of the web user concerning the experience. In our opinion this intentional dimension of the customer e-loyalty can be thought of as a customer's intention to stay in the relationship with a website provider by visiting the website and using the services that it offers.

We hypothesize that a user's sociodemographic profile can be an important factor that influences the time spent by the user on the web. For instance, students and young people could be generally expected to allocate more time for online presence than breadwinning adults for the sake of fun and entertainment. As a consequence, also time spent on particular websites could be expected to vary across web users with different profiles. A number of studies suggest that demographics can indeed act as moderators between e-satisfaction and customer e-loyalty [Anderson and Srinivasan, 2003; Simon, 2001].

The sociodemographics can also play an important role in the formation of the feelings and opinions towards websites [Ranaweera, 2003]. Important website features could in this respect include the design of the website, look and feel of the pages, or style and amount of graphics. It can be expected, as an example, that younger web surfers could favour fancy animations or music add-ons more than older users who would be irritated with the page overload. The technology in which the website is created and presented, being either simple and familiar HTML or trendy Flash technology can also affect the perception of navigation patterns.

Besides the sociodemographics, also the user perceptions about various facets of the website quality should have an impact on the overall user experience and the actual long-term user behaviour. These facets can be different depending on the type of content that the website delivers but generally can include concepts like image, ease of navigation, trust, security, etc.

In conclusion, we will assume that e-loyalty can be modelled within a framework, in which the sociodemographic profile can act as a potential determinant both of the user perceptions of the website quality as well as of the actual loyalty behaviour, or they can play the role of variables moderating the relationship between these two constructs. Next, we will treat ease of navigation, perception of layout, and look and feel of pages as attributes of the website quality, and we will assume that the website quality can be an antecedent of e-loyalty. Furthermore, though we haven't addressed the concept of attitude in this review, we find it interesting to introduce the user attitude as the study of e-loyalty as a concept that can mediate the link between website attributes and e-loyalty [Dick and Basu, 1995]. We stress here that we do not hypothesise presence or absence of particular relationships among constructs; what we do assume is only a linear ordering of these constructs in terms of the potential causal influence. Details on this ordering and on precise variables included in the study are presented in Section 4.3.

3.3.2. Conclusions for the study on traditional CS&L phenomenon

In this section, we conclude with the findings on the CS&L literature review that are applicable for the case study in Chapter 5. In this study we compare several competing models of customer satisfaction and loyalty in the traditional service setting. These models are different in some theoretical hypotheses of presence, absence or direction of influence between concepts; these hypotheses are likely on the basis of our review. The models concern four theoretical constructs: Customer Satisfaction, Trust, Involvement and Loyalty.

In all the considered models we assume that customer loyalty is the ultimate dependent concept in the domain, because of its value in as a proxy for profitability [Reichheld and Sasser, 1990; Fornell *et al.*, 1996], therefore we test it only as a child of the remaining constructs.

Based on several studies [Beatty *et al.* 1988], we find that Loyalty is determined directly by Involvement. For instance, Dick and Basu [1994] posited that "the higher the involvement in a consumption category, the greater the likelihood of loyalty towards specific offerings within this category."

Morgan and Hunt [1994] demonstrate a negative relationship between Trust and propensity to leave. Anderson and Weitz [1989] have found evidence that Trust is key to maintaining continuity in conventional channel relationships. Similarly, Doney and Cannon [1997] found that Trust of the supplier firm and of the salesperson increase a buyer's anticipated future interaction with the supplier. So, Trust is a likely antecedent of Loyalty. Furthermore, Involvement and Trust are the consequences of Satisfaction.

There is a debate concerning the causal ordering between Satisfaction and Trust [Geyskens *et al.*, 1998]. Therefore Loyalty can be a consequence of Satisfaction but indirectly through Involvement and Trust.

Also, one of the building blocks of engagement is Trust [Smith and Rutigliano, 2003], hence there can exist a link from Trust to Involvement, which is conceptually close to Engagement. However, it is likely that no direct link exists between Involvement and Trust, and, additionally also no link between Satisfaction and Involvement.

Finally, there exist findings that Involvement leads to Trust rather than the other way around [Teichert and Rost, 2003].

3.3.3. Conclusions for the practical CS study

The effect of product/service attribute performance on customer satisfaction cannot be overestimated. For instance, a recent report by the market research agency Mobius [Mobius, 2002] questioned 300 American consumers of age 18 and over to determine how they rate customer service. More than one-half (54%) of the respondents said that if they are put on hold for more than 5 minutes while speaking on the phone with customer service, the service is poor. The same report states that as much as 30% of the respondents dropped service of phone companies, a problem known as churning, due to what they deemed to be bad customer service. This example shows how important it is to measure and control the performance and satisfaction with service features (attributes.) Presumably, satisfaction studies in this example could prevent on time from customers to defect by pinpointing weak aspects of customer experience with customer service, and with service time as a very important feature of customer service at this company.

Consequently, one of the primary tasks in practical customer satisfaction studies carried out by companies and other organisations pertains to determining product/service factors driving satisfaction and/or dissatisfaction [Oliver, 1996; Hill and Alexander, 1996]. The managerial results of such a study should identify possible factors as priorities for improvement to focus company resources on these factors that require better performance on the one hand, and to decrease resources on those that possibly do not have a link with satisfaction on the other hand [Naumann and Giel, 1995]. In other words, the findings of such a study should provide insight as to the importance of product/service dimensions in terms of the strength of their influence on overall (dis)satisfaction and the character of this influence.

There are a lot of empirical studies that confirm the influence of attribute level performance on satisfaction [Bearden and Teel, 1983; Bolton and Drew, 1991; Mittal *et al.*, 1998; Oliva *et al.*, 1992; Oliver, 1993; Spreng *et al.*, 1996]. Findings from a number of studies suggest that the relations between service dimension/feature performance and overall satisfaction can often be non-linear and not straightforward. For example, Mittal *et al.* [1998] investigated the link between attributes and overall satisfaction and found that attribute-level performance impacts satisfaction differently based on whether consumer expectations were positively or negatively disconfirmed. In their study, overall satisfaction was found to be sensitive to changes in low attribute levels, whereas at high levels of attribute performance, overall satisfaction showed diminished sensitivity.

More recently, Anderson and Mittal [2000] identified three main attribute satisfaction – overall satisfaction patterns: 1) *'satisfaction enhancing,'* in which an attribute shows increasing returns on consumer satisfaction (equivalent to the attractive previous attribute); 2) *'linear and symmetric,'* in which an attribute is linearly related to satisfaction, and 3*) 'satisfaction maintaining,'* in which an attribute shows decreasing returns on both satisfaction and dissatisfaction.

Motivated with these results, in Chapters 6 and 7, we develop a methodology that will help identify the character of service attributes and dimensions. More precisely, this character can be identified by means of a classification scheme in terms of driving overall satisfaction and/or dissatisfaction. We approach these links probabilistically and express probability at each level of overall satisfaction in terms of probability of satisfactory feature performance. Low, medium, and high satisfaction can each be measured with a separate function. Furthermore, in line with the findings that stated importance is often found confounding and misleading [Oliver 1997], with the presented approach we are able to derive the importance indirectly from the survey responses. According to Allen and Willburn [2002], derived importance represents "a cornerstone of customer satisfaction research."

We lean towards the conceptualisation of customer satisfaction as an evaluative process. According to this paradigm, customer satisfaction should be operationalized by measuring customer expectations, product/service features' performance, and degree of discrepancy between expectations and perceived performance. Nonetheless, it must be stressed that some authors signal that the effects of expectations, disconfirmation and performance on satisfaction may differ for durable and nondurable products. In consequence, measurement of expectations is pointless, because the whole effect of expectations is absorbed by (dis)confirmation. For example, Churchill and Suprenant [1982] indicate that it is not necessary to measure expectations because of this absorption.

Although it is widely agreed in satisfaction literature that satisfaction and dissatisfaction are conceptually different notions, since they differ in their antecedents and consequences, for notational convenience we will assume in this chapter that dissatisfaction is theoretically identical to satisfaction. Especially, we will presume that satisfaction with service dimensions can cause overall satisfaction and dissatisfaction alike. We will therefore refer to low levels of satisfaction as dissatisfaction.

3.3.4. Conclusions on presented modelling techniques

On the basis of the short review of the dominant modelling approaches used nowadays in CS&L research, the most important conclusion that we can draw is that, despite their long history, these presented approaches still have significant shortcomings and pose numerous problems to researchers. On the other hand, the known assumptions and the principles involved with the use of Bayesian networks, as presented in Chapter 2, make it that Bayesian networks can also be used for similar purposes of testing theoretical hypotheses and models.

In the light of these findings, we find it imperative that Bayesian network approach should be examined in more detail as alternative to these aforementioned techniques. So, let us see the Bayesian network modelling in action!

4. Case study 1: Inductive development of CS&L theory with BNs

4.1. Introduction

The inductive research in the strict sense starts typically with making observations about the world and recording them as data; next, the data are rearranged and analysed so as to "bring order out of chaos"; lastly, lawlike generalizations or patterns are induced [McGarry, 1936]. In our implementation we induce generalizations by developing an overall model of e-loyalty.

We could depict the process of developing a theory graphically (see Figure 4.1.1). The process starts usually with a description that includes what is the focus of the theory, what the possible facets are that play role in the phenomenon. The subsequent efforts are typically directed towards ability to predict outcomes of various actions and conditioning variables.

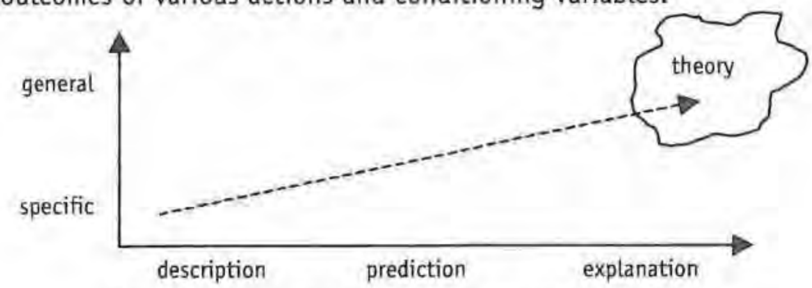


Figure 4.1.1 Schematic illustration of developing a theory.

The first attempts in the scientific search of a theory have usually narrow scope and are limited to a specific problem at hand, and concern only a particular context or a company. They might be seen as *the market research* rather than *the marketing research*. The objective of marketing research is thus at the same time to move from specific problems to more general conclusions, hypotheses and laws. With time, we should develop a scientific understanding of the phenomenon, i.e. a theory. By juxtaposing and comparing the implication of one model with those of another, and by tracing the differences to the different assumptions in the various models, we can develop a theory about phenomena under consideration. Various modelling approaches can be taken to realize each of these targets.

The ultimate goal in the process is the explanation. Hunt argues that the explanation of phenomena is the *sine qua non* condition of science [Hunt, 1991]. Hempel [1965] suggests that scientific explanations should be viewed as scientific answers to "why" questions. After Hunt [1991], we will accept that an explanatory model is "any generalized procedure or structure which purports to represent how phenomena are scientifically explained". He proposes four main normative criteria for evaluating the adequacy of purportedly explanatory structures: 1) pragmatism, 2) intersubjective certifiability, 3) empirical contents and empirical testability, but 4) first of all they must show that the phenomenon to be explained was somehow expected to occur [Hunt, 1991].

4.1.1. Objectives

The study is aimed at investigating four research questions sketched in Section 1.3. Not all these questions are covered entirely in this case study; in particular, we will attempt to answer the following research questions and achieve the following sub-objectives on the example of the e-satisfaction and loyalty research:

1. How can marketing theories be discovered by means of Bayesian networks?
    a. Examine Bayesian networks in terms of the inductive research.
    b. Discuss and illustrate specific issues in explanation of a theory:
        i. the ability for modelling of moderating effects,
        ii. the issue of accounting for mediating variables.

2. How can purported marketing theories discovered with Bayesian networks be scientifically justified (validated)?
    a. Investigate the descriptive, predictive and explanatory power of Bayesian networks.

3. What is the added value of modelling marketing problems with Bayesian networks?
    a. Demonstrate the ability of performing probabilistic reasoning (forward, backward, inter-causal) in the domain,
    b. Show the potential of performing what-if simulations.
    c. Illustrate the potential of combination of prior knowledge with data at hand.

4. What are the strengths and weaknesses of Bayesian networks in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc.

Let us discuss all these objectives in more detail.

The first question addressed in the case study in this chapter is "how can marketing theories be discovered by means of Bayesian networks?" The goal here is to examine the potential of Bayesian networks for discovery of marketing theories taking the inductive approach to research. We will not follow the strict inductivist route literally but we will adapt it so that it fits in the Bayesian network framework, taking as the application field the customer e-loyalty phenomenon. In our implementation of the inductive research, we start with making observations and recording data on online visitor bases of four different web portals; next, for each data set we develop a specific theoretical model; and lastly, we arrange the findings from these four different models into one overall model of e-loyalty. We will evaluate the results by an attempt to find similar findings in the existing literature on e-satisfaction and loyalty.

The second question addressed here is "how can purported marketing theories discovered with BNs be scientifically justified (validated)?" To be more precise, we investigate whether the outcome of this inductive process, i.e., the purported *theoretical* model of e-loyalty, satisfies the criteria of being a scientifically justified theory. In other words we ask ourselves how do we know whether the outcome is really scientific knowledge. In order to answer this question, we take the view of Hunt [1991] in that we require of a theoretical model to contribute to the understanding of the e-loyalty phenomenon by its ability of description, prediction, and explanation of this phenomenon. As regards the explanation of e-loyalty, we would like to find out for example *why* some web users are loyal, or *why* some users have favourable attitudes towards the website, etc. In answering these questions we will sketch profiles of users, rather than particular users individually. To examine the explanatory power of our Bayesian network e-loyalty model in a more systematic way, we will evaluate the purported explanatory potential of this model with the criteria revealed above as recommended by the modern empirical orientation in the philosophy of science [Hunt, 1991], namely 1) first of all, it must show that the phenomenon to be explained was somehow expected to occur, 2) be intersubjectively certifiable, 3) have empirical contents, and 4) be pragmatic. We will show that the phenomenon to be explained was indeed, by means of these models, somehow expected to occur. Theoretical models should be subject to intersubjective certifiability, so we will also evaluate our e-loyalty model in this respect. We will address the issue of empirical contents and testability, and show how our Bayesian network e-loyalty model can be empirically tested. As regards the fourth criterion for explanatory models, i.e. pragmatism, Hunt [1991] does not mention how to verify it; in our opinion, this criterion is difficult to verify; nevertheless, we will assess its pragmatism by showing the ease of model specification, or ease of use and interpretation. Other capabilities such as probabilistic inference, and what-if simulations can also be viewed as symptoms of pragmatism. Last but not least, to test the adequateness of the prediction we will make use of the models as predictive systems, and assess the predictive accuracy by comparing with other standard methods.

Important requirements of techniques aiming to contribute to the scientific understanding of marketing phenomena, and e-loyalty in particular, are the issues of moderating effects and mediating variables [Bagozzi, 1994a]. We will discuss and evaluate the capabilities of Bayesian network modelling in this context too.

The next important question to be considered in this chapter is "what is the added value of modelling marketing problems with Bayesian networks?" If we claim that the Bayesian network can fulfil the supply-demand gap of marketing modelling, we have to explore the added value of the Bayesian network approach for modelling marketing problems. To this end, we will demonstrate the ability of performing probabilistic reasoning and belief updating in the domain, and more specifically the potential of forward, backward, and inter-causal reasoning. Furthermore, we will show the potential of "what-if" simulations. As the next objective referring to the third research question, we will illustrate the potential of combination of prior knowledge with data at hand.

Last but not least, simultaneously with the flow of discussion we will also, whenever possible, explore what are the strengths and weaknesses of Bayesian networks in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc.

We stress that although we attempt to acquire the scientific understanding of the behavioural and intentional loyalty of web users towards the use of specific web sites as such, the most emphasis is, notwithstanding, put on the presentation and evaluation of our methodology, rather than on contributing to the theory of e-customer loyalty.

The remainder of this chapter has the following structure. In Section 4.2 we present the details on Bayesian network approach for inductive research. Data collection, pre-processing and operationalization of constructs are addressed in Section 4.3. Section 4.4 contains results of the application of Bayesian networks in inductive research, and discussion of other objectives. We conclude in Section 4.5, in which we also draw implications for CS&L researchers and practitioners, and provide limitations of the presented research.

4.2. Bayesian network approach for model generation

In this study, we are concerned with the use of Bayesian network technology for obtaining theoretical insights from real data by taking the model generating approach. In the model generating approach, the process of finding the best model is iterative. At each step of the procedure, the model is refined so that the next candidate model is better according to a certain criterion. To narrow the number of possible alterations, the analyst can make use of two main sources of indications [Bollen, 1989]. Usually, one takes the empirical aspect so as to optimise some statistical quality of the model, e.g., maximize its fit to the data. However, relying on the data alone can sometimes lead to nonsensical respecifications. The revision of the models can therefore also be based on the theory and substantive expertise of the modeller.

Generating theoretical models takes usually the form of an interactive process, in which the modeller plays the active role by deciding whether to accept a change in the respecified model, or by giving the direction of possible changes.

In the following case study of the model building procedure, a model constructor uses his expertise by putting some restrictions on possible structures. We will describe these restrictions thoroughly in the next section. Later, the respecification completes automatically without the interaction with the modeller and is entirely determined by the data and a search algorithm. Naturally, if the modeller is uncertain about the initial parameters, the process can and should be repeated with different initial parameters. Finally, we note also that the two sources of model's evolution, i.e., the expert's knowledge and empirical data, can be combined in the course of the model construction.

4.2.1. Experimental design

As an underlying assumption, we will presume that the observed data have been generated by an *a priori* unknown process that can be represented with some Bayesian network model. In other words, we assume that the data we observe have been generated by this unknown process. This process concerns the way that describes the e-loyalty phenomenon. Consequently, the aim of the structural learning with the Bayesian network methodology is to try to recover this process by inferring the best structure, in the form of a Bayesian network, from the observed data. In the course of the model generation many models will be evaluated with a scoring metric. The scoring metric that we will use to find the best structure has the property that given sufficient data it scores the generative model, i.e., the model from which the data could be sampled, higher than any other model that is not equivalent to the generative model [Heckerman, 1999]. So we should end up with a model that we hope is the best model that could possibly generate the observed data at hand.

As we have stated earlier in this work, learning of the Bayesian network model of any domain involves selection of a good network structure and estimation of the model's probabilistic parameters. In this study, we have taken the Bayesian approach to learning the Bayesian network model of the domain under consideration. Refer to Section 4.3 on other, non-Bayesian approaches. The Bayesian approach is suited both for the selection of the best-fitting network structure (structural learning) and estimation of the model's parameters (parameter learning) from data. First, data are used to choose the network structure with the largest posterior probability. Recall from Section 2.6.1 that the posterior probability of a candidate Bayesian network structure B_s can be according to the Bayes' rule expressed as

$$p(B_s \mid D) = \frac{p(B_s)p(D \mid B_s)}{p(D)}, \tag{4.1}$$

where $p(B_s)$ is the prior probability of the model structure B_s, $p(D)$ is the probability of the observed data D, and $p(D|B_s)$ is the likelihood of the data D given the Bayesian network structure B_s. We can obtain the probability of the data $p(D)$ by summing up the nominators in Formula 4.1 for each possible Bayesian network structure B_{si}:

$$p(D) = \sum_i p(B_{s_i}) p(D \mid B_{s_i}), \tag{4.2}$$

It follows from Equality 4.1 that if the prior probabilities $p(B_s)$ of the candidate hypothetical models B_s are equal, i.e., the models are *a priori* uniformly distributed, then the posterior probability of the model $p(B_s|D)$ is uniquely identified by the likelihood $p(D|B_s)$ of the data given the model B_s. We can ignore the probability $p(D)$ of the data, since this value is constant and independent of any particular model B_s. The likelihood $p(D|B_s)$ can be considered as the complete likelihood of the model structure B_s obtained by summing the likelihoods for all the possible instantiations of probabilistic parameters ?in the model. This can be expressed as an integral over all the possible instantiations of probabilistic parameter values θ contained in CPT's:?

$$p(D \mid B_s) = \int p(D \mid B_s, \theta) d\theta, \tag{4.3}$$

The probability of data $p(D|B_s)$ given the model as showed in Relation 4.3 is referred to as the *marginal likelihood of data D* given a Bayesian network structure B_s to denote that all the parameter values have been marginalized out of the model. For us, the most important consequence of the marginal likelihood in Relation 4.3 given model B_s is that the higher this likelihood is, the higher the chance that the model B_s has generated these data.

So, the idea of the Bayesian approach to learning Bayesian network models is that if the prior probabilities of any two candidate Bayesian network models are equal, the choice between these models boils down to the selection of the model for which the marginal likelihood of data is the largest. Under some conditions (see Section 2.6.1), the calculation of the marginal likelihood can be efficiently carried out by exploiting the property of its factorisation:

$$p(D \mid B_s) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}, \tag{4.4}$$

where n is the number of nodes in the model B_s, q_i is the number of possible instantiations of parents of the node i, r_i is the number of states of the node i, N_{ijk} are the observed counts of variable i in state k given the jth configuration of the node's parents, $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$, α_{ijk} is the prior precision, or the counts given by an expert as a priori estimates, $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$, and $\Gamma(\ )$ is the Gamma function. This score is often called the CH, or BD (Bayesian Dirichlet) score [Cooper and Herskovits, 1992].

Thanks to the factorisation in Function 4.4 for the nodes, the total score of the marginal likelihood can be split into the evaluation of the local parent-child dependencies for each node i. Let $g(x_i, p_i)$ be the local contribution that node x_i and its the parents' set p_i have to the total marginal likelihood in 4.4. Then, it results from Equation 4.4, that $g(x_i, p_i)$ can be computed using the following formula:

$$g(x_i,\pi_i)=\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+N_{ij})}\prod_{k=1}^{r_i}\frac{\Gamma(\alpha_{ijk}+N_{ijk})}{\Gamma(\alpha_{ijk})}, \tag{4.5}$$

where all the symbols have the same meaning as above. We let the prior precision α_{ijk} for each *i*, *j*, and *k* account to 1, which can be considered an uninformative prior. We have expected that this low and constant value of prior precision will not influence the posterior probability. Later, when we analyse the results, we can review the probabilities $g(x_i, p_i)$ to conclude about the relative strengths of dependencies between a node and its potential parents.

The factorisation of the BD score suggests a search in the space of possible Bayesian network structures by consideration and scoring possible combinations of parents for each node. However, the enumeration of every possible combination would be in practice infeasible. Therefore, a possible way around would be to use a search algorithm that takes only a subset of possibly highly scored parent sets into consideration. With this end in view, we applied the greedy search algorithm known as K2 [Cooper and Herskovits, 1992]. This method requires an ordering of the variables as one of its arguments. The rationale behind the ordering is that the variables that come sooner in the order are considered during the search as potential child nodes for the variables that follow next in the ordering. The selection of the greedy search algorithm was preferred over other alternatives, because it yields good results and is computationally more efficient than the edge-inversion and the exhaustive enumeration [Chickering *et al.*, 1995]. Refer to Section 2.4 for details on the greedy search algorithm.

Once a good network structure has been found, the following step consists of the estimation of the probabilistic parameters, i.e, the conditional and unconditional probability distributions $?_{ij}$ in the model. The prior precision of 1 has also the effect that none of the conditional probabilities will be 0.

Subsequently, we have dealt with the missing entries in each of the samples. In order to deal with missing entries in the data Bayesware Discoverer implements the *Bound and Collapse* (BC) method [Ramoni and Sebastiani, 1998]. This method, unlike EM algorithm or Gibbs sampling, is deterministic and the intuition behind it is that a database is able to induce bounds on parameter estimates. This method first bounds the set of possible estimates in agreement with the available data by inducing the maximum and minimum Bayesian estimates that would be obtained from all possible completions of the database. Then the resulting intervals are collapsed into a point estimate via a convex combination of the extreme points with weights depending on the assumed pattern of missing data. The empirical comparisons of this method with the EM algorithm and Gibbs sampling showed a substantial equivalence of the estimates provided by these three methods [Ramoni and Sebastiani, 1999]. The BC method makes it possible to use the Bayesian Dirichlet score when the data are missing, so that the use of approximate scores, as BIC or CS metrics, is not necessary.

4.3. Data issues

Since our research in this study deals with online audience, we have decided to rely on a database collected online. Because these data reflect a novel approach to data collection, we will describe the data collection procedure and the data characteristics in more detail next.

4.3.1. Data collection

The data that we use in the study come from a database created by means of a package called OpinionBar.[1] Web users can freely download the package from the web and install it on their computers whether at home, at work or at school. The general idea of the software is that it acts as a layer between the web and the web browser and thus monitors behaviour of the user while browsing the web and specific websites.

The information collected by the software can be categorised in three groups, namely sociodemographic data, behaviour data, and opinion data. The three categories jointly can be regarded as a rich and novel source of knowledge about the online web user experience. The data concerning the socio-demographic profile, including age, gender, education, place of residence etc., is gathered during the installation. The user is then also asked several questions about her/his hobbies, socio-economic status, familiarity with new technologies, frequency of Internet use, etc. For each piece of information the user is given some bonus points that later can be exchanged with money on a timely basis.

As regards the behavioural data, each web page (more specifically, each frame that makes up a page) accessed by the user and thus displayed in the browser, along with its full domain and URL address, query string data, as well as date, exact time and duration of viewing is registered locally by this software in a database on the user's computer. Since the software works on a very low system level, it is able to record the real time during which a web browser application has a focus on user's computer, so that even the time that the user spends on other applications running at the same time on the user's computer, e.g., word editing, is taken account of and subtracted from the total time elapsed since the user entered the website. This effective time during which the user is busy with visiting is the variable we take to calculate stickiness.

Each time that the user opens a new web browser session, the software connects to the central server and checks for updates on new surveys. Once there is a new survey for a particular website, the program performs downloading a set of questions and a precise URL address at which the survey should be activated by the program. Occasionally, when the user is in the process of visiting a web

[1] The OpinionBar software is available for downloading from http://www.opinionbar.com.

site, whose URL address matches one of those addresses on which the program keeps a survey, the survey gets activated and the user is invited to complete the survey regarding user's opinions and judgments. The user is asked to answer the questions usually about 10-15 seconds after they arrive at the website. Completed participation in the opinion polls is also rewarded with bonus points. We can thus say that both the opinion and behaviour data are recorded dynamically, i.e. while a user is actively browsing a web site. For this reason these data are unlike data used in most of other loyalty studies in which responses have a historical character. Here the data reflect the actual perception and opinion of users during the service delivery, since both occur at the same pint in time, and as such are devoid of time bias.

Both the opinion and behaviour data collected by OpinionBar during a current session are transferred to the central database server each time the user begins a new session. From the database, consisting of about 70 relations, we have first extracted the most important descriptive statistics. For example, we found out that there were about 50.000 unique OpinionBar users registered worldwide. The behaviour data were captured in 2001, during the period of which 65292 different domain name addresses were visited. The surfing behaviour data can be aggregated for each user as well as for the whole visitor base in order to provide e-metrics as the total visit duration, or frequency of visiting specific websites. Ultimately, for each website in the database we were able to extract a data set consisting of web user id's described with a range of 1) sociodemographic variables, 2) these users' opinions about the website, and 3) data reflecting the user's online behaviour at the website. Following our definition of e-loyalty, we do not take the behaviour at other sites into account, although the original data would make it possible.

The websites ranged over many types, including among others news, portal, financial, e-commerce, and adult sites. From these types of websites, we have decided to consider portal sites in further analysis for two reasons. Firstly, there was relatively more data on portals than on other website types. Secondly, portals constitute a relatively homogeneous group. Thirdly, we hoped that the e-loyalty phenomenon could be observed and, at least partly, explained on the basis of the opinion variables present in the database at hand. From the database we have extracted data describing four of the most often visited in the year 2000 portal sites in the Netherlands: WorldOnLine (www.worldonline.nl), MSN (www.msn.nl), Ilse (www.ilse.nl), and Freeler (www.freeler.nl). These sites were considered some of the main entry points to the web for Dutch users (source: verbal correspondence). The surveying and behaviour data for this sample concern the period between September 2000 and April 2001.

As regards the representativeness of the sample for the entire Internet users population, we note that the OpinionBar sample might suffer from the self-selection bias. In our opinion, the sample might be more representative for heavy Internet users, who surf online a lot. It could also be that the sample users are more familiar with computer programs than an average Internet user, although the use and installation of OpinionBar is quite easy and does not imply that its users are web experts. We have also performed a quick comparison of the samples in the current study with the respondents from other Web-based studies that we have found in the literature in terms of the basic demographic attributes [e.g., Szymanski and Hise, 2001]. In Figure 4.4.1, we show the histogram of users registered in the OpinionBar database. As regards gender, proportion of males varied between 66.3-71.3% across the selected sites. In conclusion, we can presume that our respondent base is representative of the entire population of web surfers.

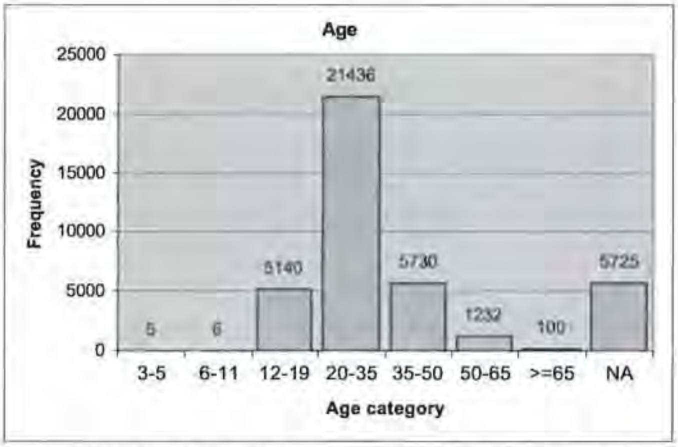


Figure 4.3.1 Histogram of age of all users registered in the OpinionBar database.

4.3.2. Measurement

Prior to the data analysis, we have performed a careful screening of the database in order to select possibly most useful attributes for our study.

Variables expressing the sociodemographic profile included for instance country, state id for US, place of residence (city), date of birth, gender, language, annual household income (< 15.000 , 15.000 - 24.000 , 25.000 - 34.000 , 35.000 - 50.000 , 50.000 >), date of account creation, last update, average internet session time, marital status (single, married, life partner), total number of persons in the household (1, 2, ..., 8+), number of children under 18 (0, 1, ..., 4+), interests & hobbies, age and gender of oldest persons in household, position in the household (breadwinner, partner of bread winner, child of bread winner, other), self-employed or work for a firm (yes, no), highest education level (high school, college, high school graduate, graduate school, college graduate, MBA, self educated), occupation, occupational industry. There were also many question concerning the use and style of the internet use, as internet use frequency (every day, more than once a week, one time per week, one time per two weeks, one time per month, less than one time per month), internet use for business or for personal (business, personal, both), connection info (home, work, school, cafe, friends house, other), primary place of use (home, work, school, cafe, friends house, other), number of users on computer at home (1, 2, 3, 3+), purpose for using the internet most (email, surfing, chatting, information, online orders, other).

To name some of the opinion variables present in the original database we can mention up-to-dateness of information, depth of information, reliability, likelihood to purchase product, opinion on assortment, importance of ease of use website, etc.

From the set of variables at our disposal, we started with deleting variables having more than 50% of missing values, which was for instance the case for income and marital status. We note that though the lack of responses for income is not surprising, the reason why there are so many missing data about marital status is not clear. Next, the variables showing not enough variation in responses, e.g., containing more than 90% of the same values were also discarded, since they presumably do not provide any sufficient insight in the considerations. One of those variables was the frequency of the Internet use, on which vast majority of users responded with the same value, namely with "everyday use". This indicates some sort of sample bias in the sense that the respondents are heavy users.

The screening of the database described above resulted in a set of variables that we ultimately have taken into further investigation. We will briefly describe them. The precise formulation of the questions and a measurement scale for each of them is reported in Table 4.3.1.

The first group of variables consists of sociodemographic variables. Among them there are: age, gender, education and position in the household. Position in the household was meant to capture the family status of the web user and indicates whether he/she is financially a head of a family, or rather a consumer of financial resources. Education is a selection of the highest education level reached by the respondent, and contains seven categories: high school, college, high school graduate, graduate school, college graduate, MBA, self-educated. We did not take the bandwidth of the connection into account, since it does not affect any of the opinion variables that we included in this study.

The second group of variables that we have eventually, predominantly on the basis of data availability and the literature review in Section 3.1, taken into further consideration were three potential dimensions of the web site quality. In this context, we have examined look and feel, layout, and the ease of navigation. The responses of each of these were all recorded with one item on 5-value rating scales.

Another variable that we include is the overall rating about the website. This is probably a theoretically more complex, more equivocal, and more capacious concept than the other three mentioned above, since it tries to capture the user's overall perception of the website. As such, it can be regarded as the user attitude. The user's attitude was also measured with one item on 5-value rating scales.

Construct	Question	Scale
Sociodemographics		
Position in the household (H)	What is your position in the household?	breadwinner, partner of breadwinner, child of breadwinner, other
Education (E)	Please select the highest education level that you achieved:	high school, college, high school graduate, graduate school, college graduate, MBA, self-educated
Age (A)	What is your date of birth?	
Gender (G)	What is your sex?	male, female
Website quality		
Look and Feel (LF)	How would you rate the Look&Feel of this site?	very negative, negative, neutral, positive, very positive
Layout (L)	In my opinion, the layout of the pages on this web site is:	not clear at all, not clear, neutral, clear, very clear
Ease of navigation (N)	Can you easily find your way through this website?	not at all, not really, neutral, somewhat, highly
Attitude		
Attitude (At)	What's your overall rating of this site?	very negative, negative, neutral, positive, very positive
Loyalty		
Likelihood to return (R)	How likely are you to return to this website?	not likely at all, certainly not, neutral, likely, very likely

Table 4.3.1. Formulation of the measurement of the constructs by OpinionBar. The behavioural dimension of e-loyalty (Stickiness) was measured independently.

In line with our multidimensional theoretical definition of the e-loyalty, we have decided to include two variables that can be linked with those dimensions. These variables allow for the operationalization of the e-loyalty. They are the likelihood to return and the stickiness. The likelihood to return refers to the user's subjective level of certainty that he/she will visit the website again. The variable Stickiness was computed as a quotient of the effective time spent on the website by a user (see section on data collection) and the number of sessions at the given website during the measurement period. Such an operationalization of has the drawback that it does not take into account the number of visits, or the total time spent on the website, yet in our opinion it could be seen as an overall, one-item measure of the behavioural dimension of e-loyalty. To conclude, these two variables can be jointly regarded as the operationalization of the two dimensions - intentional and behavioural - of e-loyalty.

Our conceptual definition of e-loyalty could be criticized for taking into account the web user's loyalty only in relation to one particular website in question, in isolation from user's potential loyalty towards alternative web sites. Consequently, by our definition, a user is loyal to a website if 1) he/she spends much time on the website and 2) is willing to return to it in the future; it does not however say anything about the amount of time that the user spends on other websites, and whether he/she is willing to return to these other websites or not. It follows that, according to our definition, the user can be loyal to many websites, and furthermore, that the user is even more loyal to a website(-s) other than the one being considered. In our opinion, this aspect does not have so much influence on the interpretation of the results once we accept that the user can be by definition loyal to more than one website. What can have influence on the results of the study is that, when evaluating the time spent on a website by different users, for each user one should consider the time spent on the website with relation to the time that he/she spends on the web in total. It could be the case that one user spends 25 min. on the website in question of total 30 min. spent on the same type of websites, while another user spends 25 min. of total 300 min. spent the same type of websites; in this example the first user is surely more loyal than the second one; however our operational definition does not recognize this problem, what can bias the results. We should therefore conclude that in the framework applied in this study, the behavioural dimension of loyalty is measured in the absolute sense, i.e., implicitly assuming that there exist fixed thresholds of time for every level of behavioural loyalty, and these thresholds are the same for each user. How we calculate these thresholds is explained in the next section.

4.3.3. Data preparation and pre-processing

The datasets that we obtained required additional cleaning and pre-processing of data to allow for efficient application of our approach and to reduce the data bias in the results. Following, we have also deleted those records for which the responses to all the five soft variables (look and feel, ease of navigation, layout, overall opinion, likelihood to return) were missing. We would like to note that we do not know why these cases are missing. The reason why these cases were deleted is because they do not contribute enough information into the joint probability distribution over these variables. For the same reason, we have also discarded those cases for which most of all variables were not observed.

In particular, for MSN we have deleted records, for which data on all opinions were missing; there were 109 such records. After cleaning, 409 cases remained in the MSN dataset. There were still quite a lot of entries with missing data, for example for Navigation there were 24.2% of missing data, and for Layout 21.8% were not reported.

The WOL dataset consisted in the beginning of 251 cases, of which 169 remained after cleaning. For Return there were 29% missing values.

For Ilse we have deleted also all the cases with missing responses to all the five opinion variables. Additionally, we have discarded some 25 other cases with missing values on navigation so that the fraction of missing data for this variable was less than 50%. As a result, there were 49.3% of missing data for navigation, which is surprisingly much more than for other datasets. The cleaning and manipulation procedure resulted ultimately in the sample of 140 records for the Ilse website.

From the data on Freeler website, we have discarded cases with missing values on the five attitude variables, which resulted in a dataset of 215 cases.

The general statistics including number of missing and valid entries for the opinion variables for the resulting data samples are presented in Table 4.3.2.

a) MSN

	Navigation	Return	LookFeel	Layout	Attitude
Valid	310	365	382	320	378
Missing	99	44	27	89	31
% Missing	24.2	10.8	6.6	21.8	7.6

b) WOL

	Navigation	Return	LookFeel	Layout	Attitude
Valid	141	119	135	131	134
Missing	28	50	34	38	35
% Missing	16.6	29.6	20.1	22.5	20.7

c) Ilse

	Navigation	Return	LookFeel	Layout	Attitude
Valid	71	138	138	130	138
Missing	69	2	2	10	2
% Missing	49.3	1.4	1.4	7.1	1.4

d) Freeler

	Navigation	Return	LookFeel	Layout	Attitude
Valid	151	178	181	172	181
Missing	64	37	34	43	34
% Missing	29.8	17.2	15.8	20	15.8

Table 4.3.2. Number of valid and missing entries for the opinion variables for each dataset.

The discrete Bayesian network model approach requires that we discretise continuous variables by transforming them into discrete ones. In our datasets there were two variables that we can regard as continuous: Age and Stickiness. The variable Age has been discretised into four intervals, being {< 19; 19 – 34; 35 – 49; >49}. This discretisation scheme is in line with the categories established for age in multiple marketing studies.

To avoid the continuous number, Stickiness was discretised into four intervals for each portal independently using the equal frequency binning principle. This principle aims ideally to find such a partition that the frequencies of objects falling in each resulting interval are the same. We have chosen the equal frequency binning scheme since this approach enforces also more reliable estimates of the conditional probabilities for Stickiness given potential parent variables by contributing to the even distribution of observations over the cells in the CPT. Whether the equal frequency binning principle has any effect on the marginal likelihood of various dependencies between Stickiness and potential parent variables, is difficult to explain without simulation studies. The resulting intervals are shown in Table 4.3.3.

	Stickiness (in secs.)			
WOL	< 69	69 - 148	148 - 319	> 319
Ilse	< 79	79 - 157	157 - 258	> 258
Freeler	< 53	53 - 99	99 - 196	> 196
MSN	< 48	48 - 117	117 - 211	> 211

Table 4.3.3. The states of Stickiness after discretization for each dataset. The intervals are average duration of visit at the website (in seconds).

We can see that the most diverse average duration time can be observed for the users of WOL and Ilse: the intervals for these two sites are remarkably wider, at the average level of 106.3 and 86, than for the two other sites, i.e., Freeler and MSN (65.3 and 70.3, respectively). In general, we can conclude that visitors generally tend to stay shorter at Freeler and MSN, and longer at WOL and Ilse. These differences in Stickiness between websites are at this stage of the study rather difficult to explain. We will later try to explain why visitors at certain websites differ in their Stickiness.

To complete the discussion of the aggregation, let us take a look at the precise frequencies of cases in each category of Stickiness. These frequencies can also be regarded as the prior marginal probabilities in the model after estimation.

Freeler		WOL		MSN		Ilse	
< 53	0.242	< 69	0.249	< 48	0.249	< 79	0.250
53 - 99	0.251	69-148	0.249	48 - 117	0.249	79 - 157	0.236
99 - 196	0.242	148-319	0.243	117-211	0.249	157- 258	0.264
> 196	0.265	>319	0.260	< 211	0.252	> 258	0.250

Table 4.3.4 Resulting frequencies for Stickiness after discretization.

For the purpose of the data analysis step, the variables describing website quality, the attitude and the likelihood to return have been ultimately created in the following way. Since the distribution of these variables was highly skewed with relatively less responses on three non-favourable values, we have pooled the three states together, e.g., for Navigation the states 'not at all', 'not really', and 'neutral' were replaced with a common label 'poorly'. The aggregation of five states into three states for these variables should not influence the interpretation of the results and can be beneficial for the later use of the model and for parameterisation. As a result, the remaining variables were aggregated in a way reported in Table 4.3.5. The resulting aggregated state received for each variable thus the meaning of neutral or moderately non-favourable or strongly non-favourable attitude. The rationale behind this aggregation scheme was to optimise the size of the tables.

	1-4	5-7	8-10
Look_Feel	negative	positive	very positive
Navigation	poorly	good	very good
Layout	not clear	neutral	clear
Attitude	negative	positive	very positive
Return	unlikely	likely	very likely

Table 4.3.5 Categories of the variables after aggregation.

The frequencies of the most important variables are reported in Tables 4.3.6-8.

Attitude	Freeler	WOL	MSN	Ilse
negative	0.381	0.254	0.272	0.094
positive	0.453	0.552	0.497	0.667
very positive	0.166	0.194	0.230	0.239

Table 4.3.6 Prior marginal probabilities for Attitude.

We can see from Table 4.3.6 that least negative attitude have the visitors of Ilse – the probability that a random visitor will have the negative attitude towards this website is 0.094. On the other hand the visitors of Freeler exhibit the highest chance to rate this website on the whole negatively.

Return	Freeler	WOL	MSN	Ilse
not likely	0.270	0.160	0.137	0.051
likely	0.348	0.395	0.282	0.384
very likely	0.382	0.445	0.581	0.565

Table 4.3.7 Prior marginal probabilities for Return.

Again, from Table 4.3.7 we can see that the users of Ilse responded most favourably with respect to their likelihood to return: there is only 0.051 probability that they responded that their return to this website in the future in unlikely.

Navigation	Freeler	WOL	MSN	Ilse
poorly	0.424	0.369	0.355	0.437
somewhat	0.305	0.433	0.361	0.366
highly	0.272	0.199	0.284	0.197

Table 4.3.8 Prior marginal probabilities for Navigation.

Distribution of aggregated responses for Navigation (see Table 4.3.8) shows a similar pattern across the four datasets with lower probability of poor assessment of the easiness of navigation by MSN and WOL users.

Finally, it is interesting to address the issue of the heterogeneity of the sample. We assume in this case study that the data come from a homogeneous sample. However, in case the sample is not homogeneous, i.e., there exist different sub-groups in population that should be considered apart from each other, one can try modeling such data by means of the mixtures of Bayesian networks [e.g., Thiesson *et al.*, 1998, 1999]. In this approach, the entire model that describes the heterogeneous sample is typically also a Bayesian network, in which some nodes are responsible for the "class" membership [*idem*].

4.4. Results

The results that we present in this section have been obtained for a specific ordering of variables as an input to the model-generating procedure. The selection of ordering was inspired by making the assumption that independent sociodemographic variables, as age and gender or education, can only act as potential ancestors of other sociodemographic variables, namely position in the household. Furthermore, the sociodemographics can act as moderators or determine the perception of website quality features, including layout, look and feel, and ease of navigation. These perceptions can in turn act as determinants of the attitude towards the website. All these aforementioned variables can at last be antecedents of the e-loyalty. We have also assumed that the user's likelihood to return can have effect on the behavioural aspect of loyalty, i.e., the stickiness. To summarise, the precise ordering was the following: {Gender, Age, Education, Position_Household, Layout, Look_Feel, Navigation, Attitude, Likelihood_Return, Stickiness}, where the variable that comes first (Gender) is during the search considered a parent for all the variables that come later in the order, the second variable (Age) is considered a child of the first variable, and a parent of the subsequent variables, and so on. We have applied Bayesware Discoverer 1.0 with parameters and preferences as discussed above to obtain the results. This software package is available free at the web address http://www.bayesware.com for downloading and evaluating.

4.4.1. Qualitative analysis

The first step in the analysis consists of the examination of the results of the structural learning algorithm. The results of the structural learning procedure for all four datasets are shown in Figures 4.4.1-4. The models shown in this figure can be regarded as the most probable structures of dependencies between the variables involved in this e-loyalty study for each website given the assumptions stated above.

The relations admit the property of d-separation, which can be used to read off both marginal as well as conditional independence relations (see Section 2.3 for more details on d-separation.) Accordingly, they are also sufficient to concisely and efficiently represent the joint probability space for the variables in this e-loyalty study. We will now review briefly the dependencies resulting from the best fitting model structures for each portal site one by one, but we postpone the analysis of the strength of the links until the following sections. As a way of clarification, we need to note that the directionality of relationships was established based on the literature review in Chapter 3. Furthermore, whether a theoretical relationship has been found to exist, which is indicated by the arrow on the following figures, is established by the greedy search procedure in the space of possible models, as described in Section 2.5.1.3.

1) *WOL*

The most likely structure of dependencies for the WOL dataset is displayed in Fig. 4.4.1.

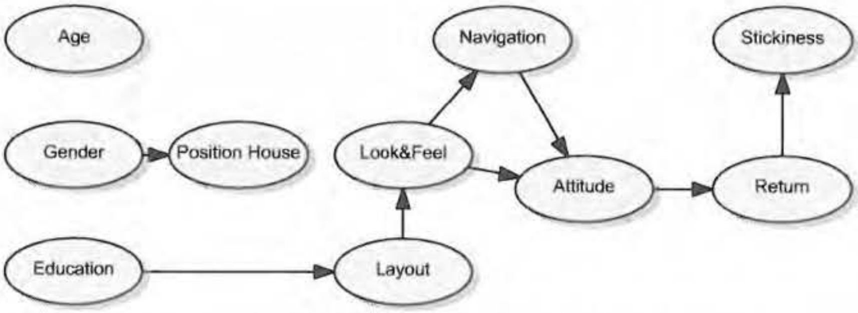


Figure 4.4.1. The most likely model structure found for the WOL data.

The joint probability of all the variables in the model $p(All)$ can be in line with the structure in Figure 4.4.1 expressed as:

$$p(All) = p(G)\, p(A)\, p(E)\, p(PH|G)\, p(L|E)\, p(LF|L)\, p(N|LF)\, p(At|LF, N)\, p(R|At)\, p(S|R),$$

where the letters denote the variables consistently with the symbols used in Table 4.3.1. Thanks to the finding of the most likely model structure and the factorisation above, we can represent the joint probability space for the variables much more efficiently now. In place of 217727 parameters we would now need only 75 non-redundant parameters to provide the probability of every possible instantiation of all the variables.

Let's take a look at the consequences of the found structure. The variables Gender and Pos_Household seem to be related only to each other, whereas Age seems neither to be relevant to loyalty nor to any other feature. We can see that Education is directly related with Layout. Look&Feel influences directly Navigation and Attitude. Navigation is also a determinant of Attitude. Furthermore, there is a link between Attitude and Return. In this model, Attitude can be thus regarded as a classical mediating variable, since it mediates the link between perceptions of website quality attributes and intentional measure of e-loyalty. We can see also that the variables that we have conceptualised as measures of customer e-loyalty, i.e. Return and Stickiness, are interdependent - a positive result (because as measures of one underlying concept they should ideally be somewhat correlated). The model suggests that Stickiness and other variables in the domain are independent given the value of Return, so that once the value of Return is known, our beliefs regarding Stickiness are not altered.

2) *MSN*

As regards the MSN dataset, the most likely model structure that describes the probabilistic interdependencies between the variables involved in the study is shown in Figure 4.4.2.

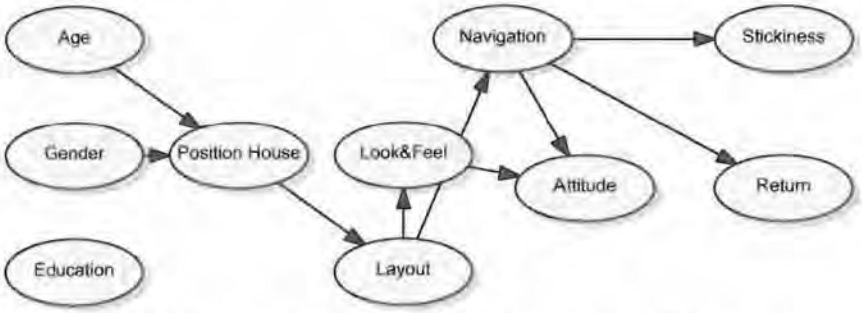


Figure 4.4.2. The most likely model structure found for the MSN data.

The dependency structure in Figure 4.4.2 can be again read as the factorisation of the joint probability distribution in the domain:

$$p(All) = p(G)\, p(A)\, p(E)\, p(PH|G, A)\, p(L|PH)\, p(LF|L)\, p(N|L)\, p(O|LF, N)\, p(R|N)\, p(S|N),$$

Age and Gender are associated with Pos_Household, whereas Education is irrelevant to the domain. Pos_Household is most likely determinant of Layout. Look&Feel and Navigation are influenced by Layout, and at the same time determine Attitude. Loyalty variables, i.e. Return and Stickiness, are influenced only by Navigation.

3) *Freeler*

The most probable Bayesian network structure of dependencies is shown in Figure beneath.

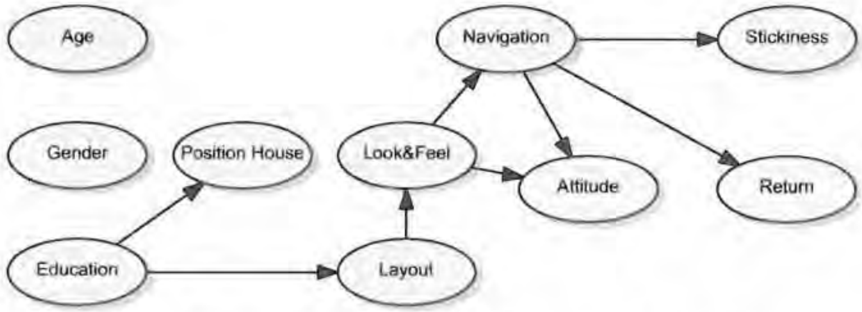


Figure 4.4.3. The most likely model structure found for the Freeler data.

According to the structure of dependencies in Figure 4.4.3 the joint probability can be factorised as

$$p(All) = p(G)\,p(A)\,p(E)\,p(PH|E)\,p(L|E)\,p(LF|L)\,p(N|LF)\,p(O|LF, N)\,p(R|N)\,p(S|N).$$

Age and Gender are irrelevant again. Education seems to determine both Pos_Household and Layout. Look&Feel influences both Attitude and Navigation and is itself determined by Layout. Navigation also relates again to Attitude. Loyalty variables, i.e. Return and Stickiness, are as is the case in MSN, influenced only by Navigation.

4) *Ilse*

The structural learning for the Ilse dataset yielded the network structure presented in Figure 4.4.4.

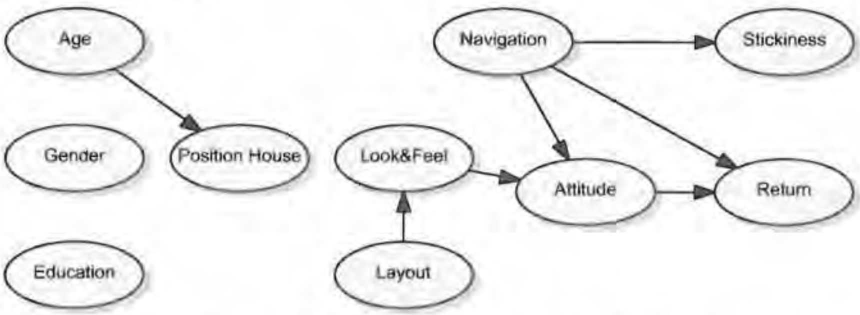


Figure 4.4.4. The most likely model structure found for the Ilse data.

The joint probability distribution can be thus factorised as in the equation below:

$$p(All) = p(G)\,p(A)\,p(E)\,p(PH|A)\,p(L)\,p(LF|L)\,p(N)\,p(O|LF, N)\,p(R|N, O)\,p(S|N).$$

Age is the only ancestor of Pos_Household, while Gender and Education are irrelevant to the rest of the variables. The socio-demographics generally are qualities not related to the website perceptions and loyalty. Look&Feel is determined by Layout, and it also determines Attitude. Attitude is besides Look&Feel also influenced by Navigation. Return is determined both by Attitude and Navigation, whereas Stickiness is determined alone by Navigation.

These above models can be regarded as causal networks under the assumption that every statistical association derives from causal interaction, and that there are no hidden common causes that could play a role in the domain [Chickering *et al.*, 1995; Heckerman *et al.*, 1997].

4.4.2. Statistical validation

In the statistical sense, the validation of the structure of the above models is based on the measure of the posterior probability of the model (see Formula 4.1 and 4.3). On the basis of this measure we can conclude that these models are valid conceptualisations of the domain in question. Furthermore, we conclude that the links are significant - the issue of the significance of links will be addressed in the next section.

Keeping in mind the constraints of the prior ordering and the greedy nature of the algorithm, we can conclude that the models outperform any other alternative model in its ability of explaining the data. To be more precise, our conclusion is that these are probably the best models in that they best explain the given data, and that other models might also provide good explanation but are less probable. We cannot however conclude *categorically* whether these models are significant in the absolute classical sense using the Bayesian approach, as we have taken.

Of course, we could attempt to calculate the posterior probability of each model as in Formula 4.1. To this end, we would have to assign prior probabilities to every possible Bayesian network structure before conducting the tests. If we were indifferent as to the quality of these structures *a priori*, then these priors could be uniform. Such an approach would require also, in addition, scoring of every other model that was excluded in the greedy search with its marginal likelihood, and summing up products of those likelihoods with structure priors to obtain the probability of observed data $p(D)$ (see Equation 4.2). Then, by Formula 4.1, we could be able to calculate the posterior probability of the network structure. The computational effort needed would be enormous, so we have abstained from this approach.

4.4.3. Likelihoods of specific links

As a form of the validation of the entire model, we can also consider the extent to which the dependencies between a node and its parent nodes hold. This can be expressed again with the posterior probability, and more precisely, with the marginal likelihood that a given node has some other nodes as parent nodes. This probability is calculated for each set of nodes that is a candidate during the search process using the Bayesian score as in Formula 4.5. This score is referred to as the Cooper-Herskovits score, or the Bayesian Dirichlet score [Cooper and Herskovits, 1992; Chickering *et al.*, 1995]. For the definition of this Bayesian score see also Section 2.6.1. We recall here that the Bayesian score is equivalent to the posterior probability of the entire model. More exactly, the posterior probability of the entire model is only proportional to the Bayesian score, since the calculation of the posterior probability requires, besides the Bayesian score, also taking account out of the probability of the data $p(D)$, and the prior probability of the model $p(B_s)$. The calculation of the likelihoods of specific links between variables is possible since the probability of the entire model factorises into the product of n factors, where n is the number of nodes in the model. For each node i we obtain hence a measure by which we can compare probabilities that the node i has certain other nodes as its parents. The parents can be regarded as potential determinants of node i in the sense of social sciences. This probability can also be used to gain some empirical insight into the potential nature of the cause-effect relationship between variables, but it must be stressed here that the causal interpretation of links is a controversial issue and is subject to a more thorough discussion. Anyway, the Bayesian score applied to local parent-child dependencies gives a good account of the validity of the model since the total marginal loglikelihood of the model is a sum of these local probabilities.

A more vivid account of the probability of a dependency between a node and a particular set of parent nodes can be provided with the Bayes factor. The Bayes factor is defined as:

$$\alpha = \frac{P(D \mid \pi_{s0})}{P(D \mid \pi_{s1})} = \frac{e^{\log P(D \mid \pi_{s0})}}{e^{\log P(D \mid \pi_{s1})}}, \tag{4.6}$$

where π_{s0} and π_{s1} are some specific sets of parents for a node in focus, and $p(D|\pi_{s0})$ is the likelihood that the node has parent's set π_{s0}.

More accurately, the Bayes factor can be seen as an indication how much confidence we can have that a node in question has a particular set of nodes as its parent nodes. It can also show how probable it is that a particular node is the only parent of a node, so that we can compare this probability to the probability that another set of nodes are potential parents. In other words, it shows how much we loose, or gain, when we adopt or reject some dependencies.

In the following tables we show the marginal loglikelihood and the Bayes factor between the dependency and other dependencies explored during the search process. In these tables, the columns contain the names of nodes that are evaluated as immediate predecessors of the node displayed in the title bar. Each dependency is uniquely identified by a set of parent variables. The numbers in the two rightmost columns are the marginal loglikelihood that the combination of nodes contained in the given row is the parent set and the Bayes factor, respectively.[1]

The following tables contain each dependency explored during the search for best parents' set for the four user databases. It is worth to note at this point the marginal likelihood measure is a measure that by its nature takes into account the complexity of the model, without any particular expression for penalizing the complexity.

We have taken a closer look at the dependencies between attitude, stickiness, likelihood to return, and layout and their immediate antecedents. These are perhaps the most interesting variables for investigation. Tables with marginal likelihoods of parents' configurations for other variables can be found in Appendix A.

Antecedents of Attitude

Let us review the Attitude towards the website and its potential determinants first. The likelihoods referring to each possible combination of determinants of this variable across the four datasets are presented in Tables 4.4.1-4. For instance, the top row in Table 4.4.1 shows the most probable set of parent nodes for Attitude, namely Look_Feel and Navigation. The marginal loglikelihood of this dependency is –112.301, and if we assume that all the legal combinations of parents are *a priori* equally likely, then the posterior probability amounts to $e^{-112.301}$ = 1.69E-49. The next ranked dependency in this table is the one including Look_Feel alone; it is less probable, since its probability equals approximately $e^{-113.117}$ = 7.48E-50. The difference is thus relatively small. Dividing these two statistics yields the Bayes factor. In the second row, we can read off the value of Bayes factor as equal 2.261. Since Look_Feel and Navigation are the most probable parents of Attitude, the Bayes factor for this combination is one, as we divide two equal numbers. The values in the remaining rows show in the same way how many times the most likely parent's set is more probable than the parent's set contained in that given row. For instance, it is approximately 207939 times more probable that the parents of Attitude are Look_Feel and Navigation than that the parents are the fourth most likely dependency, namely Look_Feel and Layout. Similarly, it is also 183505.5 and 405956 times more probable that Layout and Navigation are the single parents, respectively.

[1] As a way of clarification, we note that the order at which the variables are shown in each row has no significance; it is the set of parent nodes that has the meaning. Rows that do not contain any names, e.g., row 11 in Table 4.4.1, refer to the local structure in which a node has no parents.

Rank	Potential parents for Attitude			MLL	Bayes factor
1.	Look_Feel	Navigation		-112.301	1
2.	Look_Feel			-113.117	2.261
3.	Layout			-124.421	183505.5
4.	Look_Feel	Layout		-124.546	207939
5.	Look_Feel	Gender		-124.688	239665.6
6.	Navigation	Look_Feel	Education	-124.702	243044.5
7.	Navigation	Look_Feel	Gender	-124.981	321258.1
8.	Navigation			-125.215	405956
9.	Look_Feel	Pos_Household		-137.311	7.27E+10
10.	Look_Feel	Age		-137.434	8.22E+10
11.				-138.621	2.7E+11
12.	Education			-140.909	2.66E+12
13.	Look_Feel	Education		-143.020	2.19E+13
14.	Navigation	Look_Feel	Pos_Household	-144.048	6.13E+13
15.	Gender			-144.303	7.91E+13
16.	Navigation	Look_Feel	Layout	-145.469	2.54E+14
17.	Pos_Household			-146.235	5.46E+14
18.	Age			-150.595	4.27E+16
19.	Navigation	Look_Feel	Age	-160.756	1.11E+21

Table 4.4.1. The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Attitude and other dependencies explored for WOL data.

Rank	Potential parents for Attitude			MLL	Bayes factor
1.	Look_Feel	Navigation		-251.717	1
2.	Look_Feel	Layout		-258.382	784.463
3.	Look_Feel			-288.546	9.88E+15
4.	Navigation	Look_Feel	Gender	-294.120	2.60E+18
5.	Navigation	Look_Feel	Layout	-294.436	3.57E+18
6.	Look_Feel	Pos_Household		-296.068	1.83E+19
7.	Layout			-300.118	1.05E+21
8.	Look_Feel	Gender		-302.973	1.82E+22
9.	Navigation			-309.740	1.58E+25
10.	Navigation	Look_Feel	Pos_Household	-315.704	6.15E+27
11.	Look_Feel	Age		-328.156	1.57E+33
12.	Navigation	Look_Feel	Age	-334.443	8.46E+35
13.	Look_Feel	Education		-362.549	1.36E+48
14.	Pos_Household			-367.711	2.37E+50
15.	Navigation	Look_Feel	Education	-392.225	1.05E+61
16.				-399.504	1.52E+64
17.	Gender			-405.892	9.06E+66
18.	Education			-415.109	9.13E+70
19.	Age			-415.682	1.62E+71

Table 4.4.2. The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Attitude and other dependencies explored for MSN data.

Rank	Potential parents for Attitude			MLL	Bayes factor
1.	Look_Feel	Navigation		-116.681	1
2.	Look_Feel	Layout		-122.804	456.370
3.	Look_Feel			-138.878	4.37E+09
4.	Navigation			-139.107	5.49E+09
5.	Navigation	Look_Feel	Gender	-141.159	4.27E+10
6.	Navigation	Look_Feel	Layout	-141.573	6.46E+10
7.	Navigation	Look_Feel	Education	-143.411	4.06E+11
8.	Navigation	Look_Feel	Pos_Household	-146.527	9.16E+12
9.	Look_Feel	Gender		-154.017	1.64E+16
10.	Layout			-157.836	7.47E+17
11.	Look_Feel	Education		-160.087	7.09E+18
12.	Look_Feel	Pos_Household		-170.473	2.3E+23
13.	Navigation	Look_Feel	Age	-172.831	2.43E+24
14.	Education			-177.116	1.76E+26
15.	Look_Feel	Age		-177.298	2.12E+26
16.				-191.114	2.12E+32
17.	Gender			-196.439	4.35E+34
18.	Pos_Household			-201.561	7.29E+36
19.	Age			-204.286	1.11E+38

Table 4.4.3.The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Attitude and other dependencies explored for Freeler data.

Rank	Potential parents for Attitude			MLL	Bayes factor
1.	Navigation	Look_Feel		-64.481	1
2.	Navigation			-66.669	8.923
3.	Look_Feel	Navigation	Pos_Household	-66.943	11.731
4.	Look_Feel	Navigation	Gender	-72.627	3449.637
5.	Navigation	Gender		-75.711	75363.47
6.	Navigation	Layout		-77.778	595412.3
7.	Look_Feel	Navigation	Layout	-78.387	1095618
8.	Look_Feel	Navigation	Age	-78.507	1234488
9.	Navigation	Age		-81.213	18491323
10.	Navigation	Pos_Household		-83.330	1.54E+08
11.	Look_Feel			-96.648	9.33E+13
12.	Look_Feel	Navigation	Education	-97.381	1.94E+14
13.	Navigation	Education		-104.050	1.53E+17
14.	Layout			-109.159	2.53E+19
15.				-120.577	2.3E+24
16.	Gender			-125.434	2.96E+26
17.	Pos_Household			-126.310	7.12E+26
18.	Age			-135.047	4.43E+30
19.	Education			-147.243	8.78E+35

Table 4.4.4.The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Attitude and other dependencies explored for Ilse data.

The most probable parents for Attitude in all the four datasets are Navigation and Look_Feel. Please note that these variables are much more likely to be common parents of Attitude than one of these variables were parents individually. Also Look_Feel and Layout are collectively likely parents. There is however no doubt to come to the conclusion about the most likely parents.

We can see that Look_Feel, Navigation and Layout are in many different combinations close at the top of in each table. Conversely, the remaining variables appearing in the ordering before Attitude, i.e. the sociodemographics, situate in the second half of each table. This finding suggests that the perceptions of website features are more important than the sociodemographic profile in determining the user's Attitude. Moreover, these perceptions can only mediate the relations between sociodemographic characteristics of users and the Attitude. We note that these results are quite plausible.

Finally, we should mention also the ranks of the configuration, in which Attitude would have no immediate parents. These configurations are ranked 11^{th}, 16^{th}, 16^{th}, and 15^{th} in the list of 19 parent combinations explored. Taking into account also the Bayes factor we can state with firmness that Attitude is determined by certain variables, at least for the data at hand. To be precise, these variables are most likely Look_Feel and Navigation.

On the basis of these partial, but reasonable results, it is worthwhile to remark that the marginal likelihood scores of the dependencies between a node and its parents can be helpful in developing theoretical models. Furthermore, the measure is very intuitive.

Antecedents of Likelihood to Return

We can perform a similar analysis for the other variables as well. Let us first take a look at Return – the variable that depicts the likelihood to return to website. The results are presented in Tables 4.4.5-8.

Let us consider the most probable antecedents of Return in more detail. Navigation is twice the most likely single parent (for MSN and Freeler data), once the second most likely single parent (for Ilse), and once the fourth most likely single parent (for WOL). Moreover, it is the most likely parent commonly with Attitude in one case (for Ilse). Attitude is once the most likely and single antecedent of Return (for WOL). The definitive assessment about the most likely determinants is more difficult.[1]

In order to decide what is the most likely configuration of Return's parents on average, we could take the arithmetic mean of the ranks of the configurations. It turns out that the highest average rank is achieved by Navigation as the only parent. The mean position is 2 (ranks: 4, 1, 1, and 2) for this variable. The second best average position, as of 4 (ranks: 9, 4, 2, 1), is attained by

[1] By the way, on this example we can also see the greedy nature of the search algorithm.

Navigation and Attitude as common causes of Return. In this ranking, we haven't taken the common influence of Navigation and Look_Feel into account since this configuration does not appear in the results for the WOL dataset, though it has a high mean in the three other cases (ranks: 3, 4, 4).

Rank	Potential parents for Return		MLL	Bayes factor
1.	Attitude		-115.537	1
2.	Look_Feel		-121.142	271.770
3.	Layout		-123.521	2935.453
4.	Navigation		-126.244	44671.9
5.	Education		-126.269	45820.65
6.	Attitude	Gender	-126.644	66678.73
7.			-126.694	70051.19
8.	Gender		-130.704	3864345
9.	Attitude	Navigation	-131.346	7341251
10.	Attitude	Layout	-133.050	40334581
11.	Attitude	Look_Feel	-134.516	1.75E+08
12.	Pos_Household		-138.386	8.38E+09
13.	Age		-138.923	1.43E+10
14.	Attitude	Pos_Household	-141.276	1.51E+11
15.	Attitude	Age	-143.109	9.43E+11
16.	Attitude	Education	-144.453	3.61E+12

Table 4.4.5.The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Return and other dependencies explored for WOL data.

Rank	Potential parents for Return		MLL	Bayes factor
1.	Navigation		-280.188	1
2.	Navigation	Pos_Household	-290.619	33903.52
3.	Navigation	Look_Feel	-290.712	37210.58
4.	Navigation	Opinion	-291.778	108068.3
5.	Navigation	Gender	-292.728	279362.5
6.	Navigation	Layout	-303.207	9.93E+09
7.	Layout		-303.367	1.17E+10
8.	Navigation	Age	-320.911	4.85E+17
9.	Opinion		-324.933	2.71E+19
10.	Pos_Household		-329.058	1.68E+21
11.	Look_Feel		-330.277	5.67E+21
12.			-351.274	7.45E+30
13.	Gender		-357.898	5.61E+33
14.	Navigation	Education	-361.599	2.27E+35
15.	Age		-367.727	1.04E+38
16.	Education		-376.453	6.42E+41

Table 4.4.6.The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Return and other dependencies explored for MSN data.

Rank	Potential parents for Return		MLL	Bayes factor
1.	Navigation		-148.499	1
2.	Navigation	Opinion	-154.742	514.146
3.	Navigation	Gender	-166.316	54675957
4.	Navigation	Look_Feel	-166.370	57709249
5.	Layout		-168.282	3.9E+08
6.	Navigation	Layout	-168.684	5.84E+08
7.	Opinion		-173.537	7.48E+10
8.	Navigation	Education	-177.710	4.86E+12
9.	Look_Feel		-178.481	1.05E+13
10.	Education		-179.149	2.05E+13
11.	Navigation	Pos_Household	-182.135	4.05E+14
12.			-199.478	1.38E+22
13.	Navigation	Age	-203.680	9.22E+23
14.	Gender		-205.719	7.09E+24
15.	Pos_Household		-210.185	6.16E+26
16.	Age		-216.435	3.19E+29

Table 4.4.7.The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Return and other dependencies explored for Freeler data.

Rank	Potential parents for Return			MLL	Bayes factor
1.	Navigation	Opinion		-77.158	1
2.	Navigation			-78.355	3.307
3.	Navigation	Look_Feel		-83.316	472.017
4.	Navigation	Gender		-87.580	33571.87
5.	Opinion	Navigation	Gender	-89.519	233457.4
6.	Opinion	Navigation	Look_Feel	-92.676	5483271
7.	Navigation	Age		-96.626	2.85E+08
8.	Opinion	Navigation	Layout	-97.635	7.82E+08
9.	Navigation	Layout		-97.898	1.02E+09
10.	Opinion	Navigation	Pos_Household	-98.521	1.89E+09
11.	Navigation	Pos_Household		-102.145	7.1E+10
12.	Opinion	Navigation	Age	-106.303	4.54E+12
13.	Opinion	Navigation	Education	-109,467	1.08E+14
14.	Look_Feel			-110.986	4.91E+14
15.	Opinion			-111.430	7.65E+14
16.	Pos_Household			-120.797	8.95E+18
17.				-121.391	1.62E+19
18.	Layout			-121.749	2.32E+19
19.	Navigation	Education		-123.738	1.7E+20
20.	Gender			-124.508	3.66E+20
21.	Age			-135.514	2.2E+25
22.	Education			-145.954	7.54E+29

Table 4.4.8.The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Return and other dependencies explored for Ilse data.

The average-rank approach does not though take into account the relative likelihood of the parents' configuration. This can be improved by considering also the Bayes factor.[1] To find out which parents' configuration is more probable in general, we can simply multiply the relevant Bayes factors. For instance, the relative likelihood of the link from Navigation alone versus the link from Navigation and Attitude can be calculated as the product (7341251/44671.9) * 108068.3 * 514.146 * (1/3.307) = 2761126059, so it is 2761126059 times more likely that Navigation is the only parent than Navigation and Attitude are common parents. Similarly, it is (1/44671.9) * 2,71E+19 * 7,48E+10 * 2,31361E+14 = 1,04995E+40 times more likely that Navigation is the only parent than Attitude alone is the single parent of Return.

In conclusion, we can say that the most probable parents for Return seem to be Navigation. We note again that these results can be easily accepted intuitively. As a matter of fact, recent research in drivers of e-loyalty shows that ease of navigation is one of the most important website characteristics that could contribute to e-loyalty [e.g., Gommans *et al.*, 2001; Zeithaml *et al.*, 2002].

Another finding that we found very interesting is the high position of Gender in combination with other user perception in the rankings. Especially in combination with Navigation, Gender seems to be an important parent of Return. This pair of variables is ranked 5th, 3rd, and 4th in MSN, Freeler, and Ilse data respectively. This finding seems interesting also on theoretical grounds since it suggests a moderating effect of Gender on the relationship between Navigation and Return. We come back to this issue later in the text, as for the true meaning of this result we have to consult the conditional probability tables of Return.

Antecedents of Stickiness (behavioural dimension of e-Loyalty)

Now we will examine possible determinants of the other dimension of the e-loyalty, namely the average duration of visit at a website, viz. Stickiness. Tables 4.4.9-12 contain the probabilities of different combinations of parents of Stickiness.

As regards Stickiness, in three out of four times, Navigation is the single most probable parent of this construct. Only in case of the WOL dataset, Return is ranked first, and Navigation is on the 11th position. Return, especially in configuration with Navigation, seems to be the second most likely parent.

To make sure that Navigation is the most probable parent of Stickiness, we checked the relative likelihood by taking the marginal likelihood scores again from all datasets. We found out that the likelihood is 4,06E+72 times bigger than the probability of Return being the determinant.

[1] We cannot base this decision on the average marginal likelihood score (MLL) because the number of cases varies between the datasets.

Rank	Potential parents for Stickiness		MLL	Bayes factor
1.	Return		-184.082	1
2.	Attitude		-205.156	1.42E+09
3.	Layout		-205.813	2.74E+09
4.	Return	Layout	-207.271	1.18E+10
5.	Return	Gender	-207.934	2.28E+10
6.	Return	Education	-208.911	6.07E+10
7.	Return	Look_Feel	-212.825	3.04E+12
8.	Look_Feel		-212.956	3.46E+12
9.	Return	Navigation	-214.736	2.06E+13
10.	Return	Attitude	-215.964	7.01E+13
11.	Navigation		-221.294	1.45E+16
12.	Return	Age	-223.575	1.42E+17
13.	Education		-228.532	2.02E+19
14.	Return	Pos_Household	-232.822	1.47E+21
15.			-242.932	3.62E+25
16.	Gender		-250.437	6.57E+28
17.	Pos_Household		-253.929	2.16E+30
18.	Age		-263.756	4E+34

Table 4.4.9.The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Stickiness and other dependencies explored for WOL data.

Rank	Potential parents for Stickiness		MLL	Bayes factor
1.	Navigation		-458.495	1
2.	Layout		-468.596	24353.92
3.	Navigation	Layout	-472.876	1759786
4.	Navigation	Return	-477.976	2.88E+08
5.	Navigation	Gender	-488.056	6.89E+12
6.	Navigation	Look_Feel	-492.822	8.09E+14
7.	Navigation	Opinion	-496.727	4.02E+16
8.	Navigation	Pos_Household	-503.690	4.24E+19
9.	Return		-530.619	2.1E+31
10.	Pos_Household		-537.916	3.1E+34
11.	Navigation	Age	-540.159	2.93E+35
12.	Opinion		-554.724	6.19E+41
13.	Look_Feel		-558.866	3.9E+43
14.			-577.020	2.98E+51
15.	Gender		-587.315	8.83E+55
16.	Age		-603.428	8.78E+62
17.	Education		-615.519	1.57E+68
18.	Navigation	Education	-626.967	1.47E+73

Table 4.4.10.The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Stickiness and other dependencies explored for WOL data.

Rank	Potential parents for Stickiness		MLL	Bayes factor
1.	Navigation		-232.942	1
2.	Navigation	Return	-252.752	4.01E+08
3.	Navigation	Gender	-259.517	3.48E+11
4.	Navigation	Look_Feel	-261.551	2.66E+12
5.	Navigation	Layout	-261.605	2.81E+12
6.	Layout		-262.481	6.74E+12
7.	Navigation	Opinion	-266.330	3.16E+14
8.	Return		-272.575	1.63E+17
9.	Navigation	Pos_Household	-272.586	1.65E+17
10.	Education		-274.282	8.99E+17
11.	Opinion		-275.077	1.99E+18
12.	Look_Feel		-275.758	3.93E+18
13.	Navigation	Education	-276.220	6.25E+18
14.			-306.964	1.4E+32
15.	Gender		-313.767	1.26E+35
16.	Pos_Household		-320.914	1.61E+38
17.	Navigation	Age	-322.361	6.82E+38
18.	Age		-334.195	9.41E+43

Table 4.4.11.The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Stickiness and other dependencies explored for Freeler data.

Rank	Potential parents for Stickiness		MLL	Bayes factor
1.	Navigation		-119.338	1
2.	Navigation	Gender	-131.992	313002.1
3.	Navigation	Opinion	-139.748	7.31E+08
4.	Navigation	Return	-143.025	1.94E+10
5.	Navigation	Age	-145.901	3.44E+11
6.	Navigation	Pos_Household	-147.343	1.45E+12
7.	Navigation	Layout	-147.448	1.62E+12
8.	Navigation	Look_Feel	-149.756	1.62E+13
9.	Navigation	Education	-168.669	2.66E+21
10.			-202.389	1.17E+36
11.	Layout		-204.524	9.91E+36
12.	Pos_Household		-206.668	8.45E+37
13.	Gender		-211.711	1.31E+40
14.	Return		-211.983	1.72E+40
15.	Look_Feel		-213.330	6.61E+40
16.	Opinion		-215.299	4.74E+41
17.	Age		-224.222	3.55E+45
18.	Education		-242.704	3.78E+53

Table 4.4.12.The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Stickiness and other dependencies explored for Ilse data.

Antecedents of Layout

We have decided to gain more insight in the potential determinants of Layout because on the basis of the structural learning results, we found it surprising that some of the sociodemographic profile can play the role of an important factor, probably a moderator, in determining perception of the web page layout.

Based on the results in Tables 4.4.13-16, we can see that Education is the most likely Layout's parent in two datasets, namely Freeler and WOL. We should note that the second most likely parents' set for Layout seems to be a set that does not contain any variables, i.e., the empty set, at least by taking only the rank into consideration. Indeed, taking into account the Bayes factors achieved through the four datasets it turns out that it is 1,99E+10 more likely that Layout has no parents than the parent would be Education.

Rank	**Potential parents for Layout**		**MLL**	**Bayes factor**
1.	Education		-135.026	1
2.			-140.890	352.129
3.	Gender		-145.262	27889.35
4.	Pos_Household		-149.917	2931427
5.	Age		-153.339	89791422
6.	Education	Gender	-154.428	2.67E+08
7.	Education	Pos_Household	-168.827	4.78E+14
8.	Education	Age	-178.173	5.48E+18

Table 4.4.13. The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Layout and other dependencies explored for WOL data.

Rank	**Potential parents for Layout**		**MLL**	**Bayes factor**
1.	Pos_Household		-308.652	1
2.	Pos_Household	Gender	-330.484	3.03E+09
3.			-334.838	2.36E+11
4.	Gender		-341.154	1.31E+14
5.	Pos_Household	Age	-345.823	1.39E+16
6.	Age		-349.984	8.92E+17
7.	Education		-360.071	2.14E+22
8.	Pos_Household	Education	-428.808	1.52E+52

Table 4.4.14. The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Layout and other dependencies explored for MSN data.

Rank	**Potential parents for Layout**		**MLL**	**Bayes factor**
1.	Education		-141.568	1
2.			-164.208	6.8E+09
3.	Education	Gender	-167.244	1.42E+11
4.	Gender		-169.730	1.7E+12
5.	Pos_Household		-172.504	2.73E+13
6.	Age		-178.797	1.47E+16
7.	Education	Pos_Household	-181.677	2.63E+17
8.	Education	Age	-201.812	1.46E+26

Table 4.4.15. The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Layout and other dependencies explored for Freeler data.

Rank	Potential parents for Layout	MLL	Bayes factor
1.		-132.508	1
2.	Gender	-136.884	79.454
3.	Pos_Household	-137.042	93.114
4.	Age	-147.732	4088336
5.	Education	-159.491	5.23E+11

Table 4.4.16. The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Layout and other dependencies explored for Ilse data.

Next we have compared other different parents configurations by inspecting the Bayes factors. We have found that the none-parent scenario is 1,32E-02 less likely than the situation in which Pos_Household was the only parent. Finally, we have also compared the likelihood of the link from Gender relative to the link from Pos_Household. In three datasets, Gender is ranked higher than Pos_Household. However, the overwhelmingly high score of Pos_Household for MSN data makes it again in total 6,64E+10 more likely than as if Gender was parent of Layout.

To gain more understanding into the nature of the dependencies between Layout and Pos_Household, we should however look into the conditional probability table of this dependency.

4.4.4. Overall model of customer e-loyalty

The most likely model structures that we have found thus far were specific for each website separately. These results were obtained for data samples of small size ranging from 140 to 409 cases. In our opinion, due to the small sample size, they should rather be viewed as lacking generalization potentials for each specific portal apart, let alone for the population of Dutch portal web users in general. However, since the theoretical insight into the true e-loyalty phenomenon is not the objective in our thesis in opposition to the BN methodology, we will continue as if these sample sizes were big enough and try to generalize the findings for all the four websites in order to build an overall model of the e-loyalty. We would also like to note that the OpinionBar users might not be a good representation of web users in general, so we will implicitly restrict our analysis to this group of web users.

One way to build an overall model would be to pool all the data together, and repeat the same procedure as above again. However, the results would be then biased by the data of MSN users. Furthermore, the approach that we here apply enables drawing conclusions from different studies when no original data are available. Of course, the overall model we try to construct here is again tentative and is subject to subsequent justification and corroboration.

Let us first summarize the results of the structural learning algorithm across the four websites under consideration. To this end, we have summed up all the occurrences of the same direct dependencies that exist in at least one learned model. These counts, presented in Table 4.4.17, express how many times a particular link from a parent node to a child node is present in the four models. We have taken only the rank of the dependency into account without referring to the likelihood of the dependency. We note also that these counts have been calculated regardless whether a given node is the only parent or one of the parents. For instance, count 2 in the first row in this table denotes that Gender is the parent, or one of the parents, of Position_Household in 2 out of the 4 cases.

From	To	Counts
Gender	Pos_Household	2
Age	Pos_Household	2
Education	Pos_Household	1
Education	Layout	2
Pos_Household	Layout	1
Layout	Look&Feel	4
Layout	Navigation	1
Look&Feel	Navigation	2
Look&Feel	Attitude	4
Navigation	Attitude	4
Navigation	Return	3
Navigation	Stickiness	3
Attitude	Return	2
Return	Stickiness	1

Table 4.4.17. Counts of direct dependencies summed up across the four datasets.

The results in Table 4.4.17 suggest that sociodemographic profile of a visitor age and gender are generally irrelevant to loyalty and attitudes. The most important portal site's feature is the ease of navigation on the website as it directly affects the likelihood of return and average duration of the visit. Furthermore, we note that three links are present in all the four datasets. These links are from Layout to Look_Feel, from Look_Feel to Attitude, and from Navigation to Attitude. Three occurrences can be noticed also for two other links from Navigation to Return, and from Navigation to Stickiness. As regards the Stickiness it is clear that its parent should be Navigation. For three online user sets, this relation is most probable, and only for the users of WOL this relation is ranked 11th and is less probable than other links.

To construct a possible overall model of e-loyalty resulting from the four website specific models in question, we can compare the ranks of each dependency in the tables above for each variable or calculate the average rank. The average rank is not however an optimal measure, since it does not take the probability of each dependency into account. The probabilistic nature of these dependencies enables however the construction of the overall model consistently with our probabilistic framework. The value of the marginal loglikelihood of each parents' set, reported in Tables 4.4.1-16 in column "MLL", cannot be taken into account, as this value depends on the number of cases, which is different for each dataset, and reflects actually the probability of data given a parent configuration. We can, however, in questionable instances resort to the Bayes factor between the dependencies. The Bayes' factor can be used to compare different dependencies given that the prior probabilities of each dependency are equal.

$$\frac{p(D_1 \mid \pi_0)}{p(D_1 \mid \pi_1)} \Big/ \frac{p(D_2 \mid \pi_0)}{p(D_2 \mid \pi_1)} = \frac{p(\pi_0 \mid D_1)p(\pi_1 \mid D_2)}{p(\pi_1 \mid D_1)p(\pi_0 \mid D_2)},$$

Our findings related to the overall model are based on Tables 4.4.1-16 and the tables contained in the appendix. As an example, we will carry out a comparison between two potential parents' set for Return: {Navigation} vs. {Navigation, Attitude}. Let us take the set {Navigation} as the reference set. On the basis of the WOL dataset (Table 4.5.5), we can see that the set {Navigation} is ca. 164.35 times probable than the set {Navigation, Attitude}, because

$$\frac{e^{-126.244}}{e^{-131.346}} = e^{5.102} \approx \frac{7341251}{44671.9} \approx 164.35\,.$$

By looking at the Bayes factor in Tables 4.4.6-7 for MSN and Freeler, we see that the set {Navigation} is 108068.3 and 514.146 times more likely, respectively, than the set {Navigation, Attitude}. However, in the case of Ilse, the set {Navigation, Attitude} is 3.307 times more likely than {Navigation}; or, in other words, {Navigation} is 0.302 times as likely as {Navigation, Attitude}. In total, we have the following multiplication:

$$164.35 * 108068.3 * 514.146 * 0.302 = 2761347305.$$

It follows that it is 2761347305 times more likely that Navigation is the only parent of Return than it is a common parent with Attitude. This result could be found without performing this calculation, as we can see that in every case, Navigation alone is ranked higher than the combination Navigation with Attitude. In case of other variables, this is not so obvious, and therefore we need a presented method. For all other variables, the ultimate results of most likely antecedents are shown in Table 4.4.18 below.

Node	Parents
Gender	
Age	
Education	
Pos_Household	Gender, Age
Layout	Pos_Household
Look&Feel	Layout
Navigation	Layout
Attitude	Look_Feel, Navigation
Return	Navigation
Stickiness	Navigation

Table 4.4.18. Variables and their parents in the overall model of the e-loyalty found in the study.

We can observe that Stickiness and Return are related with each other only in case. That would suggest that our two-dimensional measure of e-loyalty fails the test of construct validity, in the sense that they do not "load" on one factor.

We found that Stickiness should be the child in the overall model of Navigation. This suggests that ease of navigation has generally an effect on the average duration of visit at the website.

Return has most likely one parent, and that is Navigation. We have found also strong support for the hypothesis that Attitude can also be a common parent of Return along with Navigation. Nevertheless, the ease of navigation should be seen as the most important determinant of the intentional dimension of the e-loyalty.

The most probable determinants of Attitude are Look_Feel and LayOut. These nodes appear in each dataset as the most probable parents. The parent of Navigation is Layout, since it is in general 4.73 times more likely than the second most likely parent, viz. Look_Feel. Look_Feel has most likely Layout as the only parent, whereas Layout is influenced only by Position_Household. The parents of Position_Household are Age and Gender. These both variables are 6.22E+04 times more likely than Age alone. We have found out that Gender, Age, and Education have most likely no parents.

As a result, the above-mentioned analyses have led us to the redesigning of the final model of dependencies in the e-loyalty domain. We present the most likely general model in Fig. 4.4.5.

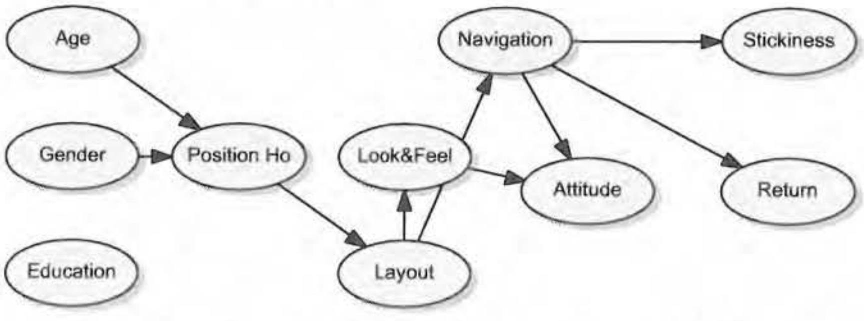


Figure 4.4.5. The most probable overall model of the e-loyalty found in the study.

Let us consider this overall model in light of the extant theory of e-loyalty. The link between ease of navigation and intention to revisit the website is supported in the literature [Loiacono *et al.*, 2000]. Chen *et al.* [2003] found support for the theoretical relation of shopping efficiency and loyalty intention, however they didn't find support for the hypothesis that website navigation is a dimension of shopping efficiency.

Consumer characteristics such as gender, age, income, are often considered as potentially having influence on customer perceptions and evaluations of service delivery [Zeithaml *et al.*, 2002; Ranaweera *et al.*, 2004], so the presence of the link between Position in the Household and the perception of Layout is theoretically sound. The indirect link between demographics such as Age and Gender and Layout through Position in the household is also very likely.

In conclusion, we find that the discovered overall model of customer e-loyalty is to a large extent consistent with the extant e-satisfaction and loyalty literature. So as to leave no doubt, we must take into account that the literature in the field of e-loyalty is still young and thus not rich in theoretical findings. For instance, we were not able to find any study that includes the overall attitude towards the website service.

4.4.5. Quantitative analysis and strength of relationships

Apart from the qualitative dimension described above, the resulting models are also parameterised in terms of probabilities that collectively constitute the model's quantitative dimension. These probabilities are assigned to the dependencies in form of conditional probabilities for variables that have parents in the model, as well as to unconditional (prior) probabilities for variables that have no parent nodes. Examination of the conditional probabilities tables is useful for the investigation of the actual strength of dependencies that exist between variables that are qualitatively related. Furthermore, the probabilities along with the presupposed qualitative dependencies describe the e-loyalty domain for each dataset apart. These probabilities have been assessed using the Bound-and-Collapse method of Ramoni and Sebastiani [1998] for dealing with missing data. The empirical comparisons of this method with the EM algorithm and Gibbs sampling showed a substantial equivalence of the estimates provided by these three methods [Ramoni and Sebastiani, 1999]. Simulation studies with data from other domains showed that the method can be used to reliably estimate the conditional probabilities even when a large part of data for some variables is missing [*idem*]. Of course, whether this characteristic of the BC method holds for the data at hand must be evaluated in further studies; here we presume that the estimation procedure makes a reliable estimation. Moreover, we decided to include cases with some variables missing, because they can provide valuable information on the relationships between variables that are observed in these cases.

We consider here only some selected conditional probability tables for the illustration. The remaining conditional tables can be found in Appendix B.

The conditional probabilities can be found in the conditional probability tables where a separate distribution over the states of a variable is held for each combination of this variable's parent states. In the tables below, the top row(s) list(s) the name of the parent variable(s) and various states that these parent variable(s) can take on. The states of the variable under consideration appear in the leftmost columns, so that the columns sum up to unity. In the row labelled "Counts", we present the observed counts of cases in each dataset that are found in the given configuration of parents' states. As the size of each data set in focus is rather small, this number should be taken into consideration to assess significance of the conditional probabilities in the configuration, so that when it is low, we should not consider the given distribution as reliable; therefore, whenever we discuss some particular conditionals of interest, then we do it by first consulting the number of observed counts.

Effects of parents on Navigation

Let us begin with Navigation, as this seems to be the most important concept that impacts loyalty. In two cases, Navigation is directly dependent on Look&Feel, and in one case on Layout.

Look&Feel	negative	positive	v.positive
Counts	50	62	21
poorly	0.528	0.382	0.202
good	0.334	0.302	0.285
very good	0.138	0.316	0.513

Table 4.4.19 Conditional probabilities for Navigation for Freeler.

Look&Feel	negative	positive	v.positive
Counts	29	64	23
poorly	0.468	0.397	0.253
good	0.461	0.432	0.406
very good	0.071	0.170	0.341

Table 4.4.20 Conditional probabilities for Navigation for WOL.

Although both the web users of Freeler and WOL, whose probabilities are contained in Tables 4.4.19-20 respectively, tend to evaluate the ease of navigation depending on the look and feel they differ somewhat in terms of the nature of this dependency. In general, the better the perception of look and feel the easier the navigation at these websites. The web users of Freeler tend, with probability 0.513, to find the navigation at this website as highly easy given that they positively evaluate the look and feel, whereas the users of WOL do it only with 0.341 probability. The users of WOL tend with quite high probability, oscillating around 0.43, i.e., from 0.406 to 0.461, to assess the navigation as somewhat easy disregarding the perception of look and feel.

The dependency between Look&Feel and Navigation can also be expressed with the Goodman and Kruskal Gamma factor - a measure that can be used to assess the direction and strength of the association between two ordinal variables [Agresti, 1984]. The Gamma factor is an ordinal statistic which is computed by using the ordinal statistical operations of "greater than", "less than" and "equal to." Using these ordinal statistics each pair of data can be classified as either tied (T), concordant (P), or discordant (Q). The Gamma factor is defined as:

$$\gamma = \frac{P-Q}{P+Q}, \tag{4.7}$$

where P is the total number of concordant comparisons, and Q is the total number of discordant comparisons [Agresti, 1984]. The idea behind the Gamma factor is that the more observations there are in the upper left and bottom right part in relation to the bottom left and upper right part of the contingency table, then the higher the surplus of the positive vs. the negative association. Possible values of the gamma factor range from -1 to 1, where values close to -1 indicate very strong negative association and values close to 1 suggest strong positive association. Let us calculate this statistic for the association between Look&Feel and Navigation. The contingency table 4.4.21 contains the counts of occurrences over the two variables.

		Look&Feel		
		negative	positive	v.positive
Navigation	poorly	14	26	6
	good	13	27	9
	very good	2	11	8

Table 4.4.21 Contingency table for observed counts of Look_Feel and Navigation for WOL data.

The number of concordant comparisons P can be calculated as

$$P = 14^*(27+9+11+8)+26^*(9+8)+13^*(11+8)+27^*8 = 1675,$$

and the number of discordant comparisons amounts to

$$Q = 2^*(26+6+27+9)+11^*(6+9)+13^*(26+6)+27^*6 = 879.$$

So, the Gamma factor obtained in line with Formula 4.7 equals $\gamma=(1675-879)/(1675+879)=0.312$. We can interpret this value as a moderate positive ordinal association, i.e., as the perception in look and feel improves, so does the perception in navigation either. Additionally, Kendall's τ-b amounts to 0.193.

Layout	not clear	neutral	clear
Counts	101	133	44
poorly	0.482	0.261	0.256
good	0.331	0.424	0.337
very good	0.187	0.315	0.407

Table 4.4.22 Conditional probabilities for Navigation for MSN data.

For the users of MSN more important than the look and feel seems to be the layout of the pages. We can see that this relation is positive: the clearer the layout, the better the perception of navigation (see Table 4.4.22).

	poorly	good	very good
Navigation	0.435	0.365	0.199

Table 4.4.23 Prior unconditional probabilities for Navigation for Ilse.

As regards the visitors of Ilse, neither Look&Feel nor Layout seems to influence the navigation properties, and therefore Table 4.4.23 contains prior (unconditional) probabilities of Navigation.

Effects of parents on Attitude

The second most important factor is attitude about the website. We will consider it next shortly. We have found that in all the four cases Attitude is directly influenced by Look&Feel and Navigation. Conditional probability distributions for these dependencies are shown in Tables 4.4.24-27.

Look&Feel	negative			positive			very positive		
Navigation	poorly	good	v. good	poorly	good	v. good	poorly	good	v. good
Counts	11	6	31	17	17	6	2	2	5
negative	0.493	0.281	0.268	0.031	0.001	0.002	0.010	0.008	0.003
positive	0.504	0.714	0.724	0.879	0.913	0.994	0.439	0.014	0.507
very positive	0.003	0.005	0.008	0.090	0.086	0.004	0.551	0.978	0.490

Table 4.4.24 Conditional probabibilites for Attitude for Ilse.

Look&Feel	negative			positive			very positive		
Navigation	poorly	good	v. good	poorly	good	v. good	poorly	good	v. good
Counts	14	26	6	13	27	9	2	11	8
negative	0.782	0.461	0.018	0.149	0.073	0.179	0.333	0.004	0.128
positive	0.215	0.537	0.965	0.777	0.855	0.555	0.006	0.333	0.005
very positive	0.002	0.003	0.017	0.075	0.073	0.266	0.661	0.663	0.868

Table 4.4.25 Conditional probabibilites for Attitude for WOL.

Look&Feel	negative			positive			very positive		
Navigation	poorly	good	v. good	poorly	good	v. good	poorly	good	v. good
Counts	50	40	12	37	55	16	11	43	28
negative	0.783	0.732	0.468	0.108	0.092	0.039	0.003	0.002	0.001
positive	0.198	0.267	0.370	0.709	0.777	0.795	0.324	0.240	0.240
very positive	0.018	0.001	0.162	0.183	0.131	0.166	0.673	0.758	0.758

Table 4.4.26 Conditional probabibilites for Attitude for MSN.

Look&Feel	negative			positive			very positive		
Navigation	poorly	good	v. good	poorly	good	v. good	poorly	good	v. good
Counts	26	23	4	17	19	6	20	7	11
negative	0.850	0.771	0.580	0.342	0.053	0.002	0.009	0.006	0.003
positive	0.149	0.227	0.283	0.615	0.898	0.860	0.248	0.322	0.260
very positive	0.001	0.002	0.137	0.042	0.050	0.139	0.743	0.672	0.737

Table 4.4.27 Conditional probabibilites for Attitude for Freeler.

The situation with two parents calls for an important question that should be asked now: which one of the two antecedents, Look&Feel or Navigation, influences Attitude in a more remarkable way, or in other words, which link of the two links is stronger? One way we could answer it is by inspecting the Bayes factors between the configurations in which Attitude has single parents (see Table 4.4.1-4). We can see that in three cases (WOL, MSN, Freeler) Look&Feel is more likely single parent of Attitude than Navigation is. It would indicate that the link between Look&Feel and Attitude is stronger than the link between Navigation and Attitude.

Another approach to compare the strengths of two links for a specific data is to examine the change in the marginal probability of the dependent variable, i.e., Attitude, as a result of instantiations of the explanatory variables, i.e., LookFeel and Navigation apart from each other. In Table 4.4.28, we can find marginal probabilities for Attitude conditionally on specific findings at its parent variables for the MSN dataset. As an example, for an average user, negative Attitude towards the web site is probable as of 0.688 given the LookFeel is also negatively perceived. Moreover, very positive LookFeel of the MSN web site results in a very low probability (0.002) of negative Attitude, if nothing else is known about the navigation (except for the marginals of Navigation). On the other hand, an average visitor is much more likely (0.153 vs. 0.002) to have negative opinion given very good evaluation of the easiness of navigation. The probability of negative Attitude given poor navigation is not as high as given negative LookFeel (0.346 vs. 0.688). Other conditionals can be interpreted accordingly.

		Attitude		
Parent		negative	positive	v positive
LookFeel	negative	0.688	0.265	0.047
	positive	0.082	0.760	0.158
	very positive	0.002	0.264	0.734
Navigation	poorly	0.346	0.466	0.189
	good	0.280	0.527	0.193
	very good	0.153	0.561	0.285

Table 4.4.28 Conditional marginal probabilities of Attitude given findings at its parents for the MSN data.

From Table 4.4.28 we can visually observe that the variation of conditional distribution for Attitude given LookFeel has more variation than given Navigation. This finding is an indication that the link between LookFeel and Attitude is stronger than the link between Navigation and Attitude.

Because the quantification of direct dependencies in a Bayesian network is done by the tables with independent conditional distributions, one distribution for each combination of state of the parent variables, it is quite obvious that these distributions can express much richer nature of the relationships than just linear. For this reason, it is not possible to provide a simple linear coefficient to describe these dependencies. Instead, the strength of the impact between two variables can be assessed by means of the entropy measure [Pearl, 1988, p.321], which would express how much influence has a variable on another one.

Let us measure the uncertainty regarding the Attitude variable by means of the entropy function first. Suppose, a random variable Q has n states $q_1, q_2, \ldots, q_n$, each with (marginal) probability $p_1, p_2, \ldots, p_n$. The information value of Q is the average information value of each state, that is

$$H(Q) = -\sum_{i=1}^{n} p_i \log_2 p_i \ .$$

For instance, if Q is ternary and has a uniform multinomial distribution, the entropy $H(Q)$ amounts to 1.585; if Q involves no uncertainty, i.e., we know that Q is in a particular state for sure, the entropy is 0. In Table 4.4.29 in the first row, we show that the entropy of Attitude amounts to 1.48. In other words, it is the sensitivity of the query node, i.e. Attitude, to a finding at the query node itself.

Let us examine the mutual information and the entropy reduction of LookFeel and Navigation in relation to Attitude. The mutual information I between the query variable Q, and the finding variable F is defined as

$$I = H(Q) - H(Q|F) = \sum_{q}\sum_{f} P(q,f)\log\left(\frac{P(q,f)}{P(q)\,P(f)}\right),$$

where $H(Q)$ is the entropy of Q before any new findings. The mutual information score can vary from 0 to $H(Q)$. The entropy reduction is then a ratio of the mutual information score and the entropy of the query node itself.

	Mutual info	Entropy reduction %
Attitude	1.480	100 %
LookFeel	0.467	31.6 %
Navigation	0.025	1.69 %

Table 4.4.29 Entropy reduction of Attitude.

Table shows that mutual information between LookFeel and Attitude equals 0.467. By relating this value to the entropy of Attitude itself, we get a measure how much entropy of Attitude LookFeel is able to reduce. This reduction amounts to 0.467/1.48 = 31.6%, whereas the node Navigation can reduce the entropy of Attitude only by 1.69%. This means that knowing LookFeel reduces uncertainty around Attitude much more significantly than Navigation.

A disadvantage of the entropy measure is that it does not take ordinality of the variables into account, so it does not communicate if the relation is positive or negative. To determine whether it is positive or negative, we could resort to the Gamma factor. We should notice that the notion of entropy is known well in psychology and has been applied in numerous studies [La Cerra and Bingham, 2002; Chen, 2003]. As Chen [2003] states "since information is the reduction of entropy and all human activities are essentially entropy processes, it is natural to understand human psychology and market patterns from the viewpoint of entropy theory." In our opinion, the entropy reduction as a measure of impact of a predecessor variable on a focus variable is an interesting alternative in the CS&L research to traditional measure of correlation or linear regression. Yet another approach to examine the relative strength of predecessor variables is based on the idea of symbolic propagation [Castillo *et al.*, 1995]. We examine this capability in the context of the practical CS research in Chapter 6.

Attitude as a mediating variable

Now, we address the potential of modelling mediating variables. A mediating variable is one that explains a relation or provides a causal link between other variables [Sekaran, 1992].

Let us go back to the most likely model generated for the WOL data. In this model, Attitude can be regarded as a good example of a classical mediating variable, since it "explains" the link between perceptions of website quality attributes and intentional measure of e-loyalty. Deciding whether a variable is a mediating variable or not consists in consulting the marginal loglikelihood of different parents' set for a variable we want to explain, and for a potential intermediary variable. As we look at Table 4.4.5, it is easy to see that Attitude is the most likely single parent of Return, and from Table 4.4.1 we can infer that Look&Feel and Navigation are the most probable parents of Attitude. Such an analysis, and plausible theoretical insight is sufficient to conclude that Attitude is in this specific example an intervening variable.

What's more, we can resort to the marginal likelihood of different parents' sets to test for the consequences of omitting intervening variable.

Effects of parents on Likelihood to return

Next, let us discuss the loyalty variables. Likelihood to return is in the case of Freeler and MSN determined by the ease of navigation. We can see that for both sites this likelihood is much bigger when a user can easily find their way on the website, i.e., the return is very likely with the probabilities 0.662 and 0.758, and likely with the probabilities 0.278 and 0.195, respectively. The data on WOL show that if the Attitude is very positive than there is only 0.042 probability that the user is unlikely to visit the website again, and as much as 0.916 probability that it is very likely that they will return.

Navigation	poorly	good	v. good
Counts	53	42	37
unlikely	0.348	0.199	0.060
likely	0.360	0.409	0.278
very likely	0.292	0.392	0.662

Table 4.4.30 Conditional probabilities for Return for data on Freeler.

Navigation	poorly	good	v. good
Counts	102	106	81
unlikely	0.148	0.125	0.046
likely	0.289	0.332	0.195
very likely	0.562	0.542	0.758

Table 4.4.31 Conditional probabilities for Return for data on MSN.

Attitude	negative	positive	very positive
Counts	28	65	26
unlikely	0.392	0.109	0.042
likely	0.357	0.553	0.042
very likely	0.251	0.338	0.916

Table 4.4.32 Conditional probabilities for Return for data on WOL.

Attitude	negative			positive			very positive		
Navigation	poorly	good	v. good	poorly	good	v. good	poorly	good	v. good
Counts	7	2	1	20	19	11	3	4	2
unlikely	0.548	0.411	0.911	0.001	0.030	0.001	0.005	0.004	0.002
likely	0.175	0.574	0.057	0.571	0.566	0.252	0.010	0.209	0.014
very likely	0.277	0.015	0.033	0.428	0.403	0.747	0.986	0.787	0.983

Table 4.4.33 Prior conditional probabilites for Return for Ilse.

On one hand, the interpretation of the conditional probabilities for the visitors of the Ilse portal should be very cautious due to the low number of observed counts. In general, we would like to stress that particular values of conditional probabilities should be viewed as reliable only when samples size is big enough. On the other hand, even when the sample is not large enough, as it is the case with data in this chapter, we should note that the Bayesian network approach seems to be successful in discovering the existence theoretical relationships or a lack of thereof, which are very likely given findings in the literature. This potential should be seen as a plausible advantage of the used Bayesian score. However, to make a firm conclusion on the effectiveness of the Bayesian network approach for small sample sizes would require using simulation and test studies on sub-samples of a larger sample.

The conditional probability tables of Stickiness, and other variables can be found in appendix.

Potential moderating effect of Gender

Let us come back to the issue of conditional probability tables for Return in the light of the possible moderating effect of Gender on the relationship between Navigation and Return. We have seen before that the combination of Navigation and Gender was scored very high in three cases: in case of MSN, Freeler, and Ilse (recall Tables 4.4.5-8). Though it is much more likely that Return is dependent on Navigation alone, we will now assume for a while, as a measure of illustration, that Navigation has indeed direct effect on Return, but Gender plays the role of a moderating variable in this relationship. Support for this assertion can be found in a number of positions in the literature [e.g., Ranaweera, 2003; Simon, 2001].

By definition, a moderating effect is "a phenomenon that affects the relationship between two or more other phenomena such that the relationship changes, depending on the level of the moderating variable" [Bagozzi, 1994, p. 372]. In the Bayesian network context, let us assume that a focal variable is directly dependent on two parent variables, one of which is an explanatory variable, and the other one is a potential moderating variable. Consider first the marginal conditional probability distribution of the dependent variable given a specific level of the explanatory variable, and conditionally on one state of the moderating variable. Now, it is legitimate to speak of a moderating effect if there exists a clear difference between this considered distribution and another distribution of the explanatory variable, conditionally on another levels the moderating variable. Of course, the more difference between these distributions, the more significant the moderating effect. Furthermore, we can even study a moderating effect for different levels of the explanatory variable, and observe whether the moderating effect gets stronger or weaker. Let us show it by an example.

In Tables 4.4.34-37, we have gathered the CPT's of Return estimated from data for a configuration, in which Gender and Navigation were parents of Return in a new supposed Bayesian network model. Let us take first the distribution for the poor level of Navigation for the Freeler data in focus. We can see that the differences in probabilities of Return between males and females, in the first two columns in Table 4.4.34, are not big, and amount to 0.095, 0.045, and 0.05 in the absolute values for various levels of Return, respectively.

Navigation	poorly		good		v. good	
Gender	male	female	male	female	male	female
Counts	35	18	30	12	25	12
unlikely	0.380	0.285	0.174	0.260	0.045	0.090
likely	0.344	0.389	0.472	0.253	0.208	0.420
very likely	0.276	0.326	0.354	0.487	0.747	0.490

Table 4.4.34 CPT of Return for the Freeler data.

Navigation	poorly		good		v. good	
Gender	male	female	male	female	male	female
Counts	23	7	14	11	12	5
unlikely	0.050	0.179	0.002	0.123	0.002	0.091
likely	0.506	0.156	0.460	0.560	0.202	0.215
very likely	0.444	0.666	0.538	0.317	0.797	0.694

Table 4.4.35 CPT of Return for Ilse.

Navigation	poorly		good		v. good	
Gender	male	female	male	female	male	female
Counts	77	25	81	25	50	31
unlikely	0.146	0.155	0.163	0.002	0.036	0.064
likely	0.310	0.228	0.343	0.301	0.128	0.308
very likely	0.544	0.617	0.494	0.697	0.835	0.629

Table 4.4.36 CPT of Return for MSN data.

The absolute values of these differences can be found in Table 4.4.38. In the first row with the data, we can see for instance that the difference in the low likelihood to return between males and females given poor perception of navigation amounts to 0.095. In the lower row in this table we show the average absolute difference taken over all the three states of Return.

Navigation	poorly		good		v. good	
Gender	male	female	male	female	male	female
Counts	32	11	30	16	13	8
unlikely	0.154	0.445	0.067	0.247	0.153	0.007
likely	0.475	0.181	0.471	0.430	0.237	0.244
very likely	0.372	0.374	0.462	0.322	0.609	0.749

Table 4.4.37 CPT of Return for WOL.

Navigation		poorly	good	v. good
Return	unlikely	0.095	0.086	0.045
	likely	0.045	0.219	0.212
	very likely	0.05	0.133	0.257
Average		0.063	0.146	0.171

Table 4.4.38. Absolute differences in likelihood to return between males and females for Freeler.

It is easy to observe that the better the perception of navigation, the bigger the difference between the distributions of Return for male and female visitors – this suggests that the nature of this potential effect is complex and not straightforward. Let us therefore inspect the distribution given the perception of navigation is very good. Now, what is more important in diagnosing the moderating effect are the probabilities of different likelihoods to return. The probability that a male user is very likely to return is 0.747 versus 0.49 for a female user, which makes up discrepancy of 0.257 – a significant result. Consequently, there is much more probability for return of males very satisfied with navigation than females. Furthermore, for females, the opinion about navigation does not seem to affect their high likelihood to return (for instance, for MSN, it oscillates around 0.63), and suggests that for women, ease of navigation is not so important, as for men, in forming their intention to return. This means, to conclude, that the link between Navigation and Return is moderated by Gender. A similar result can be noticed for the Ilse and MSN data, however, a thorough inspection of the link for users of the other websites is required to form convincing findings.

In summary, the existence of this moderating effect is not entirely confirmed by the data, and so is relatively weak. This is also reflected in the result that the "causal" influence of Navigation alone is more likely than this potential moderating effect of Gender. Nevertheless, we should stress that the important issue of moderators, in the context of CS&L research, can be traced and easily studied by our Bayesian network approach. In contrast, in SEM modelling, to study moderating effects requires much more effort, as we have mentioned in Chapter 3.

4.4.6. Description and marginal probabilities

As we have mentioned, description is one of the necessary requirements of theoretical models [Hunt, 1991]. In our opinion, in the context of Bayesian networks, the descriptive power manifests itself both in the qualitative dimension and the quantitative dimension of the model. With respect to the qualitative dimension, description of the e-loyalty phenomenon involves all the names and conceptualisations of the nodes in the model, and the presence or absence of relationships among them, whereas prior marginal probabilities of each of the nodes can be viewed as the quantitative description of e-loyalty. We could consider the chain formula for the joint probability distribution also as a form of description.

Attitude	Freeler	WOL	MSN	Ilse
negative	0.388	0.227	0.267	0.090
positive	0.460	0.562	0.516	0.737
very positive	0.152	0.211	0.218	0.172

Table 4.4.39 Prior marginal probabilities for Attitude.

The algorithms for probabilistic inference report in the first place the marginal probability distributions over the states of random variables included in the Bayesian network model. These prior distributions do not principally vary from the frequencies found in the raw data after the pre-processing step that we performed in Section 4.3 on data characteristics, in which we have discussed the prior marginal probabilities. The only differences can arise as a result of approximation of the model's parameters in case of missing values in some variables. The discrepancies are thus in practice negligible and due to the space limitations we do not show these tables here.

As an example, let us review marginals of Attitude in Table 4.4.39. From the comparison between the websites, we can conclude that visitors of Ilse have the least probability of judging this site negatively. The highest probability of a negative Attitude is characteristic for the visitors of Freeler. The probability that the visitors will judge Freeler negatively is three times higher compared with the Ilse website. The audiences of MSN and WOL have similar attitudes towards these websites. The marginal probabilities of other variables are included in Appendix C.

4.4.7. Explanation

Now, by means of our Bayesian network models of e-loyalty, we will try to answer the questions that we asked at the beginning: why some web users become loyal to portal sites, why the users have a favourable attitude towards the website. In other words, we will try to find laws that can explain the phenomenon of e-loyalty. The considerations are carried out to see if and how our approach can fulfil the requisites of the scientific explanation.

To explain the e-loyalty phenomenon we should examine the relationships between the e-loyalty variables and those variables, on which the e-loyalty depends most. According to our four models, these dependent variables are Attitude (WOL), Navigation (MSN and Freeler), and Attitude as well as Navigation (Ilse).

Hunt generalizes that "most philosophers of science agree that to seek an answer as to why a phenomenon occurred is to at least show that, given some antecedent conditions, the phenomenon was somehow expected to occur" [Hunt, 1991]. That e-loyalty can be expected to occur, we can find simply in the tables that describe the conditional dependencies, reported in Tables 4.4.30-33. For instance, let us first see why some web users of WOL might be loyal. When we do not know anything about the users of WOL, we can expect that their return to this website is very likely only with 0.445 probability (see Table 4.3.7). Now, given that these users have very positive attitude towards the WOL site, their return is very likely almost with certainty, i.e., with the probability as of 0.916 (Table 4.4.30). Hence, we can say that the very positive Attitude "explains" why the users of WOL are loyal, at least in terms of their willingness to return to this website. To be more precise, we should conclude that among the visitors who claim to return to the site, there is a high proportion of visitors who have a very positive attitude about the site, but assuming temporal ordering between attitude and willingness to return, we may speak about a notion of explanation.

Let us take another example, in which e-loyalty is caused by the attitude and ease of navigation, as is the case with Ilse. Although *a priori* there is 0.565 probability that the user will declare very likely return to the website (see Table 4.4.7), if we now introduce the premises which we think contribute to e-loyalty, condition it now on their attitude and their perception of navigation, then this likelihood raises to 0.983 (Table 4.4.33).

It is evenly important to explain why the attitude of some users is more favourable than others. For this purpose, we should consult the conditional probability tables for Attitude given its immediate antecedents, for instance of the WOL users. Unsurprisingly, it turns out that the highest probability that the return is very likely is achieved when both the perceptions on look and feel as well as on the ease of navigation are most favourable. Then this probability amounts to 0.868 (see Table 4.4.25). Likewise, the potential reasons for which web users are not loyal can also be discovered and explained.

As a form of explanation, and description alike, we can also consider a special kind of probabilistic inference, known as the most probable configuration.

Most probable configuration

Let us find out what are the specific user attitudes and loyalty outcomes that most probably occur together. They can be found as the states belonging to the most probable configuration (explanation) in the network, where a configuration is a list of states of the list of all nodes in the network. Such an examination can be valuable for a marketing researcher and practitioner alike because it shows the most probable characterisation of a respondent in terms of all the involved variables. In other words, it sketches the most likely profile of a web visitor. To avoid ambiguity, we mention that finding the most probable configuration is different than reading off the state with the highest probability marginals for each variable. In particular, prior marginals reported in Tables 4.3.3-8, show the probabilities of an average visitor in terms of each variable apart, whereas the most probable configuration reports the most likely visitor taking all the variables into account (as if we had the total joint probability table over the domain).

Return	Freeler	WOL	MSN	Ilse
not likely	0.223	0.159	0.111	0.060
likely	0.353	0.401	0.279	0.409
very likely	0.425	0.440	0.610	0.532

Table 4.4.40 Prior marginal probabilities for Return.

As an example, let us see what is the most likely profile of a visitor of the WOL portal shown in Appendix E, and compare it with the prior marginal probabilities reported in Tables 4.4.39-40 and in Appendix C. As can be seen, marginally, there is the biggest chance that an average visitor is very likely to return to the WOL site, with probability 0.44. However, when we examine what is the most probable profile in terms of all the variables, it turns out that the state of Return labelled "likely" belongs to the most probable configuration, while other variables agree with the most probable states reported in Tables 4.4.39-40 and in Appendix C with marginals.

Similarly, the most probable configuration mode enables also gaining insight as to which specific judgments and evaluations occur most likely together, or which marketing actions together will most likely cause desired effects. We can also find the maximum likely configurations *a posteriori*, i.e., when evidence is entered.

It is worth mentioning that the potential of querying for the most probable configuration can be useful in finding answer artefacts. Generally, answer artefacts are responses that are given without any consideration on the part of the respondent, and as such they do not reflect the true judgment of the respondent, e.g., endorsing only the extreme (or alternatively the middle) answer categories, or the left vs. the right positions of answer scales. If we define such answer artefacts as the most frequent responses as answer artefacts, then we can easily discover them in the data using the BN framework by obtaining the likelihood of the case given the current model (the current model is the model obtained for all data, including these artefacts). If the likelihood of the case is high and is approximately equal the value of the most probable configuration, or if the values of the random variables observed for a respondent in question correspond to the most probable configuration, then such a case can be suspected as an answer artefact. Similarly, cases that are very unlikely given the model can be also regarded as answer artefacts; very unlikely combination of values for a set of variables can be traced by the procedure known as "data conflict resolution" [Jensen, 2001]. Tracing answer artefacts by means of Bayesian networks depends however on how we will define "answer artefact".

The most probable explanation can be easily discovered with the max normal propagation mode [Jensen, 2001; Dawid, 1992]. The advantage of this propagation mode is that the most probable configuration is a by-product of the sum normal mode. The user does not have to select each possible state one at a time for all the variables, perform the probability update, and observe the resulting probability of such a configuration. The complexity of such a "manual" procedure would be thus immense and intractable. With the max-propagation inference the configuration of maximum probability can be found with only one propagation run.

In summary on the explanatory potential of theoretical Bayesian network models, we argue that the conditional probability tables and can be regarded as an instance of the inductive-statistical class of explanatory models [Hunt, 1991]. In addition, it is quite obvious that the models have empirical content, and lend themselves naturally to empirical testing, what we have shown by the use of empirical data and the marginal likelihood score, so it can be agreed that the conditional probability tables, and the Bayesian network approach in general, is subject to intersubjective certifiability. The approach is highly formalized; its "language" is expressed and based on the very well-grounded and well-known laws of probability calculus. Furthermore, we have shown that the theoretical implications of our model, obtained by the applied marginal likelihood measure, and followed from the syntax of the models (conditional probabilities) are in line with the existing literature and knowledge on e-loyalty. This supports the view that, all things considered, the Bayesian network modelling fulfils the criteria of pragmatism. In conclusion, we argue that the criteria of explanatory models, i.e., empirical contents, intersubjective certifiability, pragmatism, and expectation to occur, are supported by the Bayesian network approach to theoretical modelling.

4.4.7.1. What-if analysis

What-if analysis is a highly desirable capability of marketing models [Wierenga, *et al.*, 1994, 2000]. For instance, Rust *et al.* [2000] argue that what–if analysis have been widely applied in a wide range of marketing applications [e.g., Lilien *et al.*, 1992], and claims to use these might very well be useful in analysing the impact of customer satisfaction arena.

"What-if" analysis takes usually the form of questions such as "what will happen if ...?". Usually, the objective is to simulate the impact of different marketing-mix scenarios on some outcome variables. Rust *et al.* [2000] gives a couple of examples, for instance, "if we increase service quality, how much can we increase price, if we want to keep the same market share?"

In a similar style, we can query our e-loyalty model with "what-if" questions. Although we do not have classical marketing-mix actions included in our model, we can investigate what should be the improvements in ease of navigation, look and feel, or layout if we wished the audience was a bit more eagerly willing to return. Taking for instance the WOL data, provided the likelihood to return with states "not likely", "likely", and "very likely" was distributed 0.13, 0.36, and 0.50 instead of the original distribution 0.16, 0.40, and 0.44 for each state respectively, and given the conditional probabilities stay the same, it turns out that this increase is followed by the new distribution of Navigation with probabilities 0.378, 0.432, 0.189, instead of original 0.384, 0.433, and 0.183, and by the distribution 0.217, 0.563, and 0.219 for Look&Feel, instead of prior distribution 0.227, 0.569, and 0.204, respectively. In other words, 1.04 times better perception of ease of navigation and 1.07 times better perception increase of look & feel are required to receive this higher probability that users will come back to the web site. At the same time, this means that likelihood to return is more sensitive to changes in ease of navigation than in look & feel, so it pays more to create easier navigated websites than to work on their look and feel.

In another example, we can find out what will happen if ease of navigation is very bad and look & feel very positive in terms of the likelihood to return. In this scenario, the user is very likely to return with probability 0.734; if look & feel is only positive, the chance of very likely return drops to 0.473.

4.4.8. Probabilistic Inference

The ultimate use of a Bayesian network model is performing the probabilistic inference, that is, the acquisition of conditional probabilities from the model on condition that some variables are instantiated. Let us discuss some of the most interesting scenarios.

One of possible analysis concerns the situation in which we want to assess the loyalty of the website's target group. If the target group is supposed to consist of, for instance, well educated, breadwinning web users, then we select the relevant sociodemographic states, and observe the values of the probability distribution for the loyalty variables. If loyalty turns out to be lower than average and this trend is stable it might turn out that we should redefine the target audience and advertise on some other websites. For instance, given the visitor is breadwinning college graduate, it is 1.03 times more probable than he/she will state that he/she is unlikely to visit the website again.

The probability that user will respond most positively to all opinion specific questions is 0.045. In the comparison, probability that users will respond only positively is 0.033, and that they will respond less than favourably is 0.316.

4.4.8.1. Prediction and forward inference

In opposition to explaining, the underlying assumption behind prediction is that certain laws and theories are known, and we use this knowledge to draw conclusions about the future based on the current state [Hunt, 1991].

In order to test the predictive power of the models that we have found in the previous steps we should consider them as classifiers and run the classification task. Using Bayesian network models as classifiers consists, for each case, in examining the posterior probability distribution of a selected variable, which value we aim to predict, conditionally on the values of all the remaining variables, both antecedent as well as consequent ones. Then, the predicted variable is assigned the state that received the maximum probability in the posterior distribution. Finally, the state received by such a classification procedure is compared with the actual state of the variable that exists for the case in the dataset.

First, we have evaluated predictive accuracy for all cases using also all cases to estimate and calibrate a predictive model.[1] We refer to this procedure as batch prediction. Attitude was the variable whose state we aimed to predict. The results are presented in the first row in Table 4.4.41. For instance, for Ilse we have scored accuracy of 80.4%. This approach however is often criticized for the use of the same data twice: first in estimation, and then in prediction. Consequently, the results are often not objective.

A better approach often applied would be to divide the original dataset into two disjoint sets, called training and test sets. The data in training sets are used to learn the parameters of the model, i.e., the conditional probabilities, while the test set is used actually to perform the classification by evaluating the posterior probabilities as described above. We have applied another procedure, called cross-validation, since this routine is recommended especially in situations when there is not enough data to perform reliable parameter estimation with training sets, as is the case with small datasets. Cross-validation is a procedure that splits the original single dataset into two sets of data, train- and test-sets. By using *k*-fold cross-validation the original dataset is split into *k* disjoint test-sets, for each of which the remaining data constitute a train set. The results of all the *k*-fold classification tasks are then averaged to form an overall measure of classification accuracy.

We have performed 10-fold cross-validation procedure with Attitude as the variable whose state we aim to predict. The results are shown below in Table 4.4.41. As we can see the predictive accuracy varies from 70.9% for WOL to 76.0% for Ilse. The average predictive accuracy over the four datasets equals 73.3%. Please note that this result was achieved by taking into account also all the cases on which one or more observations were missing.

[1] We have performed this test in order to be able to compare these results with other techniques to be discussed later.

	Freeler	WOL	MSN	Ilse
Accuracy*	74.3%	76.1%	70.9%	80.4%
Correct	134	97	268	105
Incorrect	47	37	110	33
Accuracy	74.0%	72.3%	70.9%	76.0%
St. Dev.	2.990	3.439	2.246	3.605
Gamma factor	0.862	0.892	0.891	0.898
Brier Score	0.445	0.480	0.448	0.384
Brier Score UP	0.621	0.588	0.625	0.506

Table 4.4.41. Predictive power of models using 10-fold cross-validation, (*except the first row-batch prediction, Brier Score UP stands for uninformed prior).

The predictive accuracy score has two drawbacks: one that it does not take the quality of the prediction into account, and, two, that it does not correct either for the probabilistic nature of the prediction. An ideal measure that would overcome these two shortcomings of the predictive accuracy would ideally assign very high, close to unity, posterior probabilities to the true predicted states, as well as very small probabilities to the other states. In this line, a prediction that gives similar posterior probabilities to all the states of the variable under prediction is not sharp, and thus not of a high quality. One of the measures that accounts for this technique is the quadratic Brier score [*e.g.*, Gaag and Renooij, 2001]. Refer to Section 2.6.4 for more details on the calculation of the Brier score.

The results of the prediction of the Attitude with the Brier score show that the quality of the predictive power of the four website-specific models is quite high. We have chosen Attitude, and not one of the loyalty variables because it is connected with other variables in the models. The best prediction is achieved for Ilse data, as of 0.384, and the worst for the WOL data, as of 0.48. Recall that the Brier score can vary from 0 to 2, and lower values stand for better performance. It becomes apparent how good these results are when we compare these scores with the predictions obtained for two other uninformed classifiers. The first classifier is a totally uninformed classifier that assigns a state to each case at random according to the uniform distribution. The quality of such a "prediction" expressed with the Brier score is for all the four datasets equal and amounts to 0.667 (not shown in Table 4.4.41). Furthermore, if we estimate the marginal probabilities for Attitude based on the training data, and use the resulting distribution to classify the test data, then we get another classification model. The averaged Brier scores obtained for this second classifier with cross-validation procedure are reported also in Table 4.4.41 in the row named "Brier Score UP". Now we can see even better the difference between our Bayesian network models of the e-loyalty when we use it for classification and other simpler classifying systems. The most significant difference in the Brier scores occurs for the MSN data and amounts to 0.177.

Finally, let us evaluate the predictive power of our Bayesian network approach in the light of the standard approach used in the marketing science to conduct classification tasks. Probably the most widely used standard method in this respect is the discriminant analysis. Discriminant analysis is a technique that finds a set of discriminant functions based on linear combinations of the predictor variables that provide the best discrimination between the classes (states) of the predicted variable. We have carried out two tests in which Attitude was again the variable to be predicted, whereas all the variables acted as independent variables. Due to the relatively small number of cases in each dataset and to missing values, we have decided to perform the leave-one-out cross-validation procedure. In leave-one-out cross validation, each case is classified by the functions derived from all cases other than that case.

In the first test we performed the classification only for those cases that were fully observed by all dependent variables. The results of this test and the number of cases used in the analysis are reported in the first row in Table 4.4.42. We can notice a remarkable difference in predictive accuracy, raging from 74.1% to 81.3%, compared to previous results in favour of the linear discriminant analysis, but we must remember that in previous experiments with Bayesian networks we let also the data on "dependent" variables be missing. The second row contains the results of the leave-one-out validation. Now we can see that that the results are much lower than in the classification showed in Table 4.4.41. However, we must take into account that the results in these two tables cannot be so directly compared, since leave-one-out classification is different than 10-fold cross-validation.[1]

	Freeler	WOL	MSN	Ilse
Excl. cases with missing values	81.3%	74.1%	74.4%	81.0%
Cross-validated	70.8%	63.0%	71.0%	69.8%
Cases used in classification	96	81	238	63
Incl. cases with missing values	72.4%	72.4%	73.3%	77.5%
Cross-validated	63.5%	56.7%	68.5%	65.9%
Cases used in classification	181	134	378	138

Table 4.4.42. Predictive accuracy obtained by discriminant analysis in cross-validation.

However, if we include also the cases on which one or more observations on dependent variables were missing, substitute these missing values with mean value, and perform the classification, then we obtain the second test. This test produced the results shown in the lower part in Table 4.4.42. We can see that the results for the scenario with all cases included in the prediction are much worse now. A look at the accuracy achieved with leave-one-out validation leads us to

[1] Unfortunately, none of the Bayesian network software packages available to us implements the function of leave-one-out validation, and performing the procedure by hand is in practice infeasible.

conclusion that the results differ somewhat from those attained with our Bayesian network models. The best score, as of 68.5%, is found for MSN data, and the worst one, as of 56.7% for WOL data. These results are much worse than those in Table 4.4.41. However, we stress again that these results should not be compared with the results obtained with BN models, since the validation procedures were different.

We can therefore conclude that a fully correct comparison of the techniques that we applied here is not possible for different reasons. Firstly, we have made use of the feature of Bayesian networks that they enable estimation and classification even when some data are missing both in the training and test sets, whereas the discriminant analysis requires that all data are observed or that missing data are substituted with mean values. Secondly, different measures of predictive accuracy were applied due to limitations of software.

In these circumstances, in our opinion, the most "fair" test would be between the setting in which the Bayesian network approach was evaluated by means of 10-fold cross-validation, and the discriminant analysis in which all cases are included with the missing values substituted by mean values. For this test, the both approaches score on average almost the same: the average accuracy for the discriminant analysis amounts to 73.9%, whereas the Bayesian network models gave on average the accuracy of 75.4%. In the setting under consideration (Attitude as a variable to be predicted) there is thus a small difference in the power of classification between the two approaches in favour of the Bayesian network approach. Based on this result, more specifically: good prediction and theoretically sound explanation, we have positive support allowing us to conclude that the e-loyalty phenomenon is indeed constructed as our models suggest.

We should stress that although the discriminant analysis is deemed the standard approach for prediction, it is rather a simply structured model that cannot be viewed as an approach to theoretical modelling in the CS&L research compared to path-like models, such as SEM. In our opinion, due to its mathematical construction discriminant analysis makes less contribution to the scientific understanding of the CS&L theory, especially because it may not be deemed an explanatory technique.

Finally, we have compared the overall accuracy of our Bayesian network models of e-loyalty against the predictions of cumulative logit models. In the appendix C we present the results of. The cumulative logit model is a more recent technique than discriminant analysis that takes into account the ordinality of the dependent variable. Again, we have let the Attitude be the dependent variable. The comparison between Bayesian networks models and cumulative logit models can be achieved by means of the Gamma factor calculated for the confusion matrix containing the numbers of true and predicted cases. The results for the cumulative logit can be found in Table 4.4.43. We can see the difference in the results in favour of the Bayesian network models, since the values of the Gamma factor in Table 4.4.41 are much higher than those in Table 4.4.43. Therefore, to conclude, we must say that the Bayesian network models for the e-loyalty data at hand are in terms of the predictive power superior to the other standard techniques used nowadays in the CS&L research.

	Freeler	WOL	MSN	Ilse
Gamma factor	0.861	0.704	0.767	0.865

Table 4.4.43 Gamma factors for cumulative logit models.

4.4.8.2. Retrodiction and backward inference

The term retrodiction denotes making inferences about the past on the basis of present observations [Ryle, 1949]. Another scientific term used for the same procedure is postdiction [Hanson, 1963].

If we assume that the perceptions of the website and the opinions precede in time the loyalty of the visitors - a reasonable assumption - then retrodiction implies that we can determine the level of opinions or the sociodemographic profile of a visitor on the basis of the observation that loyalty is high.

In the Bayesian network terminology, retrodiction can be referred to as the backward inference, i.e., inferring about antecedents based on the known values of descendent nodes. Let us illustrate the ability of retrodiction in Bayesian networks on the example of the audience of the MSN website. Perhaps one of the most interesting observations that a marketing analyst would like to do are the socio-demographic characteristics of the most loyal audience. As a matter of fact, the network structure induced from the data suggests that education, age and gender do not have any relation with the loyalty variables. We can however find out how will the loyalty affect the position in the household. In particular, if we take that the visitor is very likely to return, then it turns out that it is a little bit more likely that he/she is a breadwinner than an average visitor. This change in our belief towards the position in the household is rather minimal since its link with loyalty is mediated by the opinions and navigation, and each of the opinion is probabilistically dependent on each other, so we cannot expect big reduction in uncertainty. Similarly, we can find out what is the probabilistic distribution of specific opinions, website elements or other antecedent variables given some causally or temporarily subsequent constructs, specifically like loyalty.

By performing the backward inference, we can also test the strengths of the links. For example, we can get marginal conditional distributions for Look&Feel and Navigation given Attitude.

Tables 4.4.44-45 contain marginal probabilities of these two variables conditionally on various states of Attitude. It is easy to notice that the distribution of Look&Feel changes more rapidly than the distribution of Navigation.

Attitude		poorly	good	v. good
Navigation	poor	0.445	0.310	0.298

	good	0.393	0.384	0.333
	very good	0.161	0.306	0.368

Table 4.4.44. Marginal probabilities of Navigation conditionally on various levels of Attitude for MSN data.

Attitude		poorly	good	v. good
Look&Feel	negative	0.842	0.168	0.070
	positive	0.156	0.747	0.368
	v.positive	0.001	0.085	0.560

Table 4.4.45. Marginal probabilities of Attitude conditionally on various levels of Attitude for MSN data.

Is the ability to retrodict a necessary condition for the understanding of the theory of CS&L, as required by our criteria? The opinions vary. For instance Hanson [1963] argues positively by claiming "every prediction, if inferentially respectable, must possess a corresponding postdiction." Following Hunt [1991], our position is that the retrodiction is a desirable characteristic of any model, however neither explanation nor prediction implies necessarily the ability of retrodiction. Nevertheless, we have shown that the capability of retrodiction is one of the advantageous characteristics of our Bayesian network modelling approach.

4.4.8.3. Inter-causal reasoning

Another useful feature of our Bayesian network e-loyalty model is its ability to perform inter-causal reasoning. It can be seen as a special kind of what-if simulation.

Let us illustrate it by an example of the Ilse data set. As we know, Look&Feel and Navigation are marginally independent random variables, unless something specific is known about Attitude or Likelihood to return. Prior to any analysis, marginal probability of Look&Feel is 0.216, 0.599, and 0.185 for negative, positive, and very positive states, respectively, as shown in Table 4.4.46a). Assume we want to assess the impact of very positive Attitude on Navigation and Attitude, so we insert the evidence "very positive Attitude." In Table 4.4.46b) we can see how this evidence influenced the marginals of Look&Feel: of course, probability of negative Look&Feel fell to 0.152, while the probability of very positive judgement raised up to 0.269. Now, we learn that perception of Navigation is very good; as a result, the distribution of Look&Feel has changed again and looks like reported in Table 4.4.46c): the probability of negative Look&Feel increased again to the level of 0.166, and at the same time the probability of very good Look&Feel decreased to 0.219. We can see that after the evidence on very favourable perception of Navigation has been introduced, the marginal probability of Look&Feel is in between the prior values and the values after very positive Attitude is assumed.

a) prior

	negative	positive	v.positive
Look&Feel	0.216	0.599	0.185

b) before evidence on Navigation

	negative	positive	v.positive
Look&Feel	0.152	0.579	0.269

c) after evidence on Navigation

	negative	positive	v.positive
Look&Feel	0.166	0.615	0.219

Table 4.4.46 Marginal probabilities for Look&Feel as an illustration of inter-causal reasoning.

The effect that instantiating Navigation had on the distribution of Look&Feel in this example is also called *explaining away* [Pearl, 1988], and we can say that very good Navigation has explained away very good Attitude, so it is now more reasonable that the probability of very positive Look&Feel becomes lower. This feature is unique and can be seen as another advantage over other modelling methods in use nowadays.

4.5. Conclusions and future research

4.5.1. Conclusions

Let us reconsider the most important results accomplished in this chapter. We have performed here an investigation into capabilities of theoretical modelling by means of Bayesian networks. For the illustration of this purpose, we designed a case study, in which we strived for understanding the drivers of customer loyalty in online environments. Let us review all these objectives in more detail.

1. How can marketing theories be discovered by means of BNs?

1.a. We have examined the potential of Bayesian networks for developing theory of e-loyalty with the inductive approach.

Since customer e-loyalty is a relatively new phenomenon, we decided to design an inductive study, in which on the basis of data we aimed to discover a possible theory of this phenomenon; we departed from a position in which we were not sure what could be the relationships between theoretical constructs in the domain; what we did assume was only the ordering of variables, from the most antecedent constructs to the ultimate ones. Next, we have let a search algorithm look for the most likely model given the data in the space of different theoretical hypotheses; since we had four different data sets for visitors of different web sites, this search was performed independently for all of them. The result was very positively surprising: the learned models are very similar to each other in terms of theoretical consequences. We can thus observe that the inductive search with the Bayesian network approach makes very reliable inference from data. Hence, we conclude that the results obtained are generic, in the sense that the differences that exist in all possible aspects of each portal site considered, and most importantly web users' perception thereof, do not have any influence on the underlying theoretical model of e-loyalty. This finding suggests also that there exists an overall model of e-loyalty that is valid generally.

Therefore, based on this last finding, we ended up the procedure with constructing an overall model of e-loyalty, derived from these four single models. Even more interestingly, we found this overall model likely to be theoretically sound given the existing e-loyalty literature. Unfortunately, since the e-loyalty literature is scarce, we were not able to find that they are fully confirmed by this literature. For this reason, our contribution to the e-loyalty phenomenon is that attitude is not so much important given the website quality, especially the ease of navigation at a portal site.

From the Bayesian network modelling perspective, we must conclude that not only the greedy nature of the algorithm that searches for the most likely model, but also the marginal likelihood score itself, as a measure of goodness of fit, proved very appropriate and successful in developing theory of customer e-loyalty.

It is necessary to note here that the inductive approach we have taken in this study is in fact very close to the exploratory research in that we can easily develop a sound theory from cross-sectional response data without imposing any hypotheses that could bias the resulting theory; of course prior ordering of variables could be argued to be one of such prior information, therefore we discuss it in more detail in the section on limitations. In our opinion, we can nevertheless conclude that the Bayesian network approach is suitable for exploratory research.

All in all, we conclude that the performance of the Bayesian network approach in inductive e-loyalty research is successful and its examination is positive.

1.b. Secondly, we contributed to the Bayesian network literature by examining and discussing the specific noteworthy issues in explanation of e-loyalty theory.

1.b.i. First, we have assessed the ability for modelling of moderating effects.

In the Bayesian network context, let us assume that a focal variable is directly dependent on two parent variables, one of which is an explanatory variable, and the other one is a potential moderating variable. Consider first the marginal conditional probability distribution of the dependent variable given a specific level of the explanatory variable, and conditionally on one state of the moderating variable. Now, it is legitimate to speak of a moderating effect if there exists a clear difference between this considered distribution and another distribution of the explanatory variable, conditionally on another levels the moderating variable. Of course, the more difference between these distributions, the more significant the moderating effect. Furthermore, we can even study a moderating effect for different levels of the explanatory variable, and observe whether the moderating effect gets stronger or weaker.

In this way, we have been able to discover a theoretically likely moderating effect of Gender on the link between Ease of navigation and Likelihood to return, though we have found that this effect is less likely than the non-moderated relationship between Ease of navigation and Likelihood to return.

In conclusion, the important issue of moderators in the context of CS&L research can be successfully traced and accounted for by our Bayesian network approach. We note that the analysis of moderating effects with the SEM approach would require two or more different models and would be more difficult to reveal [Gefen *et al.*, 2000; Bagozzi and Yi, 1989].

1.b.ii. Furthermore, we have examined Bayesian networks for the ability of discovering and modelling mediating effects.

Deciding whether a variable is a mediating variable or not consists in consulting the marginal loglikelihood of different parents' set for a variable we want to explain, and for a potential intermediary variable. What's more, we can resort to the marginal likelihood of different parents' sets to test for the consequences of omitting intervening variable.

Such an analysis, and plausible theoretical insight was sufficient to conclude that user's attitude could be regarded as an example of intervening variable. We have found that the attitude can be thought of as the mediating variable as it best explains the entire influence of ease of navigation and perception of look and feel on the likelihood to return.

In conclusion, the Bayesian network approach makes it possible to explore the effects of mediating variables and moderators.

2. How can purported marketing theories discovered with BNs be scientifically justified (validated)?

We have investigated the descriptive, predictive and explanatory power of Bayesian networks. We have explored whether the models obtained with our approach fulfil the requirements of being deemed theoretical models.

In particular, we have verified the potential of the Bayesian network methodology for explanatory modelling. Most importantly, we have found that e-loyalty can be well explained with the perception of ease of navigation along with the attitude. To be more exact, the behavioural dimension of e-loyalty, i.e., the stickiness, can be explained better with the ease of navigation, whereas the intention to return to the website can be best explained both by the ease of navigation and the Attitude. We have also verified and confirmed the explanatory power of the models by four criteria: pragmatism, intersubjective certifiability, empirical contents, and by showing that the phenomenon to be explained was expected to occur. We acknowledge that it is difficult to answer what should be the necessary depth of scientific explanation; however, if we accept the weak falsifiability criterion then our Bayesian network approach can be deemed satisfactory explanation.

Subsequently, we have evaluated the models as predictive systems. We have received satisfactory predictive accuracy. Taking into account that customer e-loyalty is in general a domain, in which it is difficult to make right predictions, we can say that this result is good. Also the higher quality of prediction expressed with the Brier score allows us to conclude that the learned models perform well as classifiers. The classification accuracy is even higher than in other classification-specific techniques.

In general, we can conclude that the Bayesian network approach can be regarded as a technique that delivers scientifically valid theory.

3. We have tried to find out what is the added value of modelling marketing problems with Bayesian networks.

3.a. First, we have demonstrated the ability of performing probabilistic reasoning (forward, backward, inter-causal) in the domain. Probabilistic reasoning can be achieved by instantiating constructs to desired states as evidence, and determining the posterior distributions for some variables in focus.

3.b. Next, the end-user of a Bayesian network-based theoretical model of CS&L can perform what-if simulations. This potential is a consequence of the ability of performing various kinds of probabilistic reasoning in one analysis. Hence, we can find the marginal distribution of any construct conditional on values of its antecedents as well as its consequences in the model at the same time. Another useful type of what-if simulations is entering likelihoods for a variable instead of instantiating it; these likelihoods cause that the variable receives new marginal distributions that we can view as desired prior marginals; now, we can read off the marginals of other variables resulting in this new marginals. In this way, we can, for instance, find out what would be marginals of loyalty on average if customer perceptions were more favourable.

3.c. Another aspect of modelling with Bayesian networks that can be seen as the added value is the potential of combination of prior knowledge with data. We have been able to make use of our knowledge in defining prior ordering of variables.

The unique features of probabilistic reasoning and what-if simulations constitute the essential added value of the Bayesian network theoretical models, apart from the necessary capabilities as developing and validation of theory.

4. What are the strengths and weaknesses of Bayesian networks in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc.

In the course of discussion presented in this chapter, we have identified several areas, in which the Bayesian network approach could be advantageous or disadvantageous vis-à-vis other, standard techniques. However, since no true comparison with other techniques was made, most of our conclusions should be corroborated in the competitive setting.

We have found the following strengths of the Bayesian network modelling approach in the context of e-loyalty research: it handles well missing data and samples of small size, it offers a way of combining prior knowledge with data, it offers a method of avoiding overfitting, it gives simple output statistics, it requires no rule-of-thumbs, it is user-friendly and easy to interpret, it enables modelling one-item operationalization of constructs, and has good predictive capabilities. Let us discuss these advantages in more detail.

First, we found that Bayesian networks can handle missing data in a very sound way. Even if there are lots of missing data, i.e., if up to 50% of all cases on a specific variable are missing, the approach performs well in the sense that it yields similar theoretical model of relationships. Missing values are imputed on the basis of the entire knowledge (theory) encoded by the model.

Next, we must also address the issue of rather small data sample sizes. In our study, the data sizes varied from 140 to 409, and for each dataset, we have received similar results in terms of existence of theoretical relationships between some variables or a lack thereof. In our opinion, it is an advantage that regardless of the sample size, which in our study varied from a small dataset to a medium size dataset, we have been able to receive similar theoretically sound results. However, further investigations with larger data samples, and sub-samples are recommended to corroborate this conclusion and to test the sensitivity of the approach to varying the number of cases.

Bayesian networks enable in an easy way the combination of accumulated knowledge and data. In this case study, we have let the prior knowledge of possible causal ordering of variables be combined with the data.[1] Some authors can see this potential as an unnecessary burden for the researcher; for others, it will be rather seen as an opportunity to make use of the accumulated knowledge. It is worth noting that the issue of combination of prior knowledge with data is an issue of lively debate between proponents of the Bayesian statistics and advocates of traditional statistics. We leave this debate aside, and we state only that the combination of knowledge is one of characteristics of the Bayesian network approach.

Subsequently, the Bayesian network approach offers a principled method of avoiding overfitting. This means that the marginal likelihood score by its nature strikes a balance between the complexity of the model and the fit to the data.

[1] Also other ways of introducing prior knowledge, or prior theory, into the developing of a theory are provided by the BN methodology. We show these capabilities in the next case study.

At the moment, actually the only goodness-of-fit statistic in use is the marginal likelihood; there is no need to calculate any numerous statistics that are hard to interpret. Therefore, no rule-of-thumbs are necessary to interpret the sufficient value of the marginal likelihood.

We found that Bayesian networks are user-friendly and easy to interpret; elementary knowledge of statistics on the level of the Bayes' rule and basic theorems in probability calculus are sufficient to interpret the consequences of the model. We argue furthermore that Bayesian networks do not require any background in advanced mathematics or statistics from the researcher to construct a model; other techniques require in these respects much expertise in advanced topics such as matrix algebra, etc. We speculate that little effort is necessitated to communicate the results to non-experts and to get them acquainted with this methodology.

We found that one-item operationalization does not pose any problem to theoretical modelling with Bayesian networks, as the indicator is treated as the latent construct itself; it should be treated as an advantage, since other techniques, such as SEM modelling, often suffer from under, or over-identification in this respect; and require at least three observed variables per construct [Steenkamp and Baumgartner, 2000].

Bayesian networks manifest not only predictive capabilities, as thanks to the probabilistic reasoning it is possible to predict, or retrodict posterior marginals for any variable in the model, but also these capabilities show good prediction accuracy.

As regards weaknesses and drawbacks of the Bayesian network approach, on basis of the work in this chapter we found the requirement of aggregation of values, the requirement of determining the directionality of causal influence, and inability to undergo the categorical validation should be treated as potential weaknesses.

More specifically, a potential weakness of Bayesian networks is that the number of categories that the variables take on should typically be collapsed. The target number of categories depends on the sample size, and should conform to the rule: the smaller the sample size, the fewer categories there should be. The rationale behind this aggregation is to avoid sparse conditional probability tables, as sparse CPT's have a negative effect on computational feasibility of parametric estimation and validation (Bayesian scoring), as well as on the reliability of specific parameters in the CPT's. This requirement should be seen as a weakness, because it can lead to the loss of potentially valuable information, and can obscure the true results.

Furthermore, a drawback of the presented exploratory search of the most likely network structures is that we need to predetermine the direction of the potential causal influence at the beginning of the research design. This can be seen in a sense as a limitation of the reliability of the findings.

Another drawback of the presented methodology is its inability to undergo the categorical validation, i.e., a Bayesian network model cannot be validated, unless it is compared with alternatives. It is so because we get a posterior probability over models we consider. That means that we cannot accept the learned model in isolation from other models. We could accept the learned model if its probability is significantly higher than any other alternative model, as is the case in Bayesian model selection. In case the best model is not remarkably better than others we should not be overconfident in the model. The problem that arises is therefore how to judge if the difference between models is big enough. This decision is usually taken on a subjective basis.

4.5.2. Implications

From research presented in this case study we can draw implications both for researchers engaged in basic research on Customer Satisfaction and Loyalty, as well as for practitioners involved in applied e-loyalty modeling.

4.5.2.1. For CS&L researchers

First of all, on the basis of the results in visitors of four main portal websites in the Netherlands we can conclude that the Bayesian network approach is suitable and performs very well both in the context of discovery of e-loyalty theory and in the context of justification of the theoretical insights into this domain.

Most importantly, we have found out based on the study under consideration that Bayesian network modelling can be successfully applied both for explanatory and predictive research. This is one of the most constructive results.

A very positive aspect of our methodology is that the search procedure delivered theoretically very appealing results also in the sense that the variables, for which most nodes were tested as potential parents, ultimately occurred to be child nodes of the variables located closer in the initial search ordering. For example, in the initial ordering sociodemographic variables were located as the most antecedent variables; the search procedure has not found that these variables are likely linked with the e-loyalty variables. This concerns the two loyalty constructs, as they were following the attitude. Such a result supports the prior ordering that we have assumed.

The marginal likelihood measure avoids overfitting. We can see that the measure by its nature strikes a balance between the complexity of the model and the fit to the data. By consulting these tables we can become convinced that the marginal likelihood makes a "fair" judgment between configurations of one, two, and three parents, namely by selecting this configuration that is the most probable.

To discriminate between models one can use Bayesian scores. An important question that a marketing researcher would often like to ask is how big the difference is in the goodness of fit between alternative theoretical models.

Especially, the more variables are involved in the modelling the smaller the relative magnitude of the model probabilities. Bayesian network modelling has no other instrumentation to judge over statistical significance other than subjective opinion. When the difference between two models seems insignificant and the predominant aim is prediction one should use Bayesian averaging instead of Bayesian model selection.

Furthermore, we argue that Bayesian networks overcome some of the deficiencies of other similar modelling approaches since they are good both for explaining marketing phenomena as well as for predicting outcome variables.

In the conclusion, we can affirm that the presented inductive approach can be successfully applied with the Bayesian network methodology. More specifically, we propose that armed with the Bayesian learning one can use Bayesian networks in order to find the best fitting model in a manner similar to an exploratory study and to discriminate between models.

4.5.2.2. Implications for applied modelling and managers

First of all, we believe that the analytical capabilities of the Bayesian network approach, including the capabilities of what-if simulations and forward, backward and inter-causal probabilistic inference can prove useful for marketing managers in practice. These capabilities provide managers with the technique to predict future behaviour and to ask diagnostic "what if" questions based on assumed marketing actions.

What's even more important, even when applied marketing modellers do not possess enough theoretical insight in order to design a model for a specific problem, they can make use of the search algorithm that will determine the most likely model. In this case, they will need a set with observational and response data, of course. Positive point in this context is that they can use one-item scales and do not need to worry too much about low response rates, or missing data, whenever they need more theoretical insight, or plan conducting customer satisfaction programmes.

All these capabilities come along with the potential of developing and validation of the theory of the marketing phenomenon in focus.

All in all, this means that managers are offered a powerful technique that, on one hand, allows for introducing theoretical insight into the model and has high explanatory value, and on the other hand, a technique that has high pragmatic value for managerial practice.

Furthermore, the findings presented in this chapter provide insight into the theory of e-loyalty that has high practical value not only to applied marketing modellers but also to web marketing managers. For example, one of the most surprising results that we have found rather unexpectedly is that the general attitude towards a portal site is not as important as the perception of ease of navigation in the formation of customer e-loyalty. With this, we would like to stress the importance of easiness of navigation, especially while designing portal sites.

We have found that the joint probability distribution of the variables in the customer e-loyalty phenomenon can be best represented (is more likely) with a probabilistic dependency structure in which visitor's sociodemographic profile is not relevant with any other variable. The findings suggest that age and gender are determinants of position in the household, which is, on theoretical grounds, a plausible result. We argue that it does not make sense to segment visitors according to these attributes in other customer e-loyalty studies.

Furthermore, we have found that, unsurprisingly, visitor opinions matter to a great extent. From the three opinions on website characteristics that we considered, visitor opinion about the ease of navigation seems to be the most important one.

4.5.3. Limitations

Apart from the contributions discussed above the approach proposed here is based on a set of assumptions that should be taken into account when interpreting and implementing the results.

One of the main limitations is a requirement of a prior ordering of variables. The specification of the prior ordering can influence the results to a large extent. The results of a study by Chickering et al. [1995] suggest that the greedy algorithm that we applied is sensitive to variable ordering. Of course, we can re-validate the results by allowing for other models starting with different search orders. Then, from among all the resulting models, the best model can be chosen on the basis of its posterior probability. We haven't performed experiments with another initial orders of variables, because based on existing theory we were quite confident in the class of models that the order initially taken implied. There are various approaches to circumvent this limitation. For instance, we could use the more time-costly edge-reversal search procedure that does not require an ordering. Other efforts are directed at the selection of the ordering, for instance, Larrañaga *et al.* [1994] use genetic algorithms to obtain the best ordering of the variables. This issue can be a topic for further research.

From the perspective of the e-loyalty theory, we agree that the concept of e-loyalty operationalized by stickiness and intention to return can have some drawbacks. Namely, the behavioural aspect might not be well accounted for by our conceptualisation. Stickiness might not be an objective measure of behavioural e-loyalty, since according to our operational definition it implies that a user that has visited the site only once for a long time, is more loyal than a user that visits regularly but shorter on average. Furthermore, it might be dependent on the Internet connection speed (bandwidth) and other factors; therefore the model we developed has a limited theoretical significance, as many important concepts are left out.

The predictive power of the models in this study was tested only for one particular variable, i.e., Attitude. We acknowledge that the capability of the theoretical models to predict should be ideally tested for more variables in order to obtain more reliable judgment in this respect. Nevertheless, the results that we present here for predictions of only one variable seem reasonably promising. Moreover, the comparison that we have made between the BN approach and the discriminant analysis is not fully appropriate, since we should use the leave-one-out validation for both approaches.

A potential threat to validity of our results, especially for the fact that all four data sets yield very similar theoretical relationships is that the data sets have many missing values. For example, the dataset that describes visitors of Ilse reports as much as 49.3% of missing data on Navigation. This could potentially have a negative effect on the value of the used Bayesian score and missing data handling of Ramoni and Sebastiani [1997] in the sense that variables with many missing values could be given more likelihood as parents. Although at the first sight, this effect is quite likely given our results and should be taken into consideration, we haven't found any convincing evidence that this effect is significant; moreover, the method is believed to be robust with respect to missing values [Ramoni and Sebastiani, 1997].

For the sake of clarity, it must be noted that any Bayesian network model that is validated on data should be viewed as explanatory for the theory it models in the extent that it explains the data, and not the process or phenomenon under focus.

Last, but not least, we have considered a scenario in which all variables were operationalized with one item scale. In theoretical studies such a scenario is rather atypical, but is quite usual in commercial studies given the limitations on the questionnaire length. We acknowledge that an approach should be able to deal with multiple item scales to account for complex latent constructs. Multi-item operationalizations are also needed to determine internal consistency reliability and construct validity [e.g., Campbell, 1969].

4.5.4. Future research

Considerations in this chapter suggest a number of topics to be addressed in future research.

One of the most urgent limitations of the work in this study that should be addressed in future work is a method that makes the specification of the prior ordering of variables not necessary. Some potential methods in this respect include genetic algorithm-based search for the best ordering [Larrañaga *et al.*, 1996; Hsu *et al.*, 2002].

Another topic for further work is to analyse the impact of different schemes of category aggregation on the results of structural learning, in terms of favouring the existence of links between constructs or the lack thereof. Similarly, studies of its impact on the strength and the character of these relationships should also be undertaken. Especially, the issues of applying the equal frequency binning principle and of the optimal reduction scheme are of significant importance in this respect.

The approach contained in this study could be viewed not as a fully eligible second generation technique, since the measurement model is not an explicit part of the model. Therefore, extending the presented approach by the possibility of handling latent constructs and measurement model should be in future undertaken. This problem is actually examined as one of the main topics in the following case study.

5. Case study 2: The Bayesian network approach in deductive CS&L research

5.1. Introduction

In the previous chapter we presented an illustration of the inductivist approach in the CS&L research. In the case study in this chapter, we will take another perspective on development of marketing theories, one that resembles more the deductive research. In the deductivist approach, we start by making speculations about a theory, forming assumptions and advancing hypotheses; next, we proceed by proposing a hypothetical model, that can be empirically tested; and ultimately, we can deduce generalizations [Hunt, 1991]. As we argue in Chapter 1, in order for a theory to be empirically testable, it must also allow for making observations and measurements [Kaplan, 1964].

In this case study, we introduce the issue of latent constructs and the measurement model in the CS&L research deliberately into the modelling task. The motivation for this task is that there are various practical and philosophical principles

The theoretical constructs studied in the marketing research, such as attitudes, customer satisfaction, are typically abstract identity. The nature of the constructs is typically complex, they have many facets, are intangible. As such, they do not lend themselves to direct measurement. Furthermore, any single indicator captures only a portion of the underlying concept that it is intended to measure; it is imperfect because it cannot capture the full theoretical meaning of the underlying construct [e.g., Steenkamp and Baumgartner, 2000]. Instead, multiple-item measurement instruments ought to necessary be used to capture the entire character of the construct indirectly [*idem*]. The multiple operationalization doctrine is based on the partial interpretation philosophy that states that though specific measures are individually imperfect, collectively they are reliable and valid measure of the underlying construct.

The imperfect nature of the measurement of individual indicators is the consequence of the measurement error. Nowadays, it is a standard procedure in marketing modelling to account for the measurement error in modelling. For instance, Steenkamp and Baumgartner [2000] argue that the correspondence between constructs and their measures should be the explicit component of marketing models [*idem*].

Using multiple items scales, gives also possibility to assess the validity of the construct. It must be remembered that we must also evaluate the reliability of the measurement scale.

The last issue that we address here is finding the dimensionality (cardinality) of latent constructs, i.e., the number of states the construct takes. We assume that the latent constructs are at ordinal level of measurement. From the modelling perspective, this is an important issue, since it can have significant effect on the performance of the model and on its complexity [Elidan and Friedman, 2001]. More importantly even, it is an important matter from the point of view of the theory and practice of Customer Satisfaction and Loyalty.

In a Bayesian network model, observed indicator variables are treated as any other node in the network, whereas the construct variables are handled as hidden nodes. The term "latent construct", or "latent variable" is used especially in the social sciences; when we however approach the modelling of the CS&L phenomenon with Bayesian networks, the term "hidden node" is more appropriate and natural. For clarity, we will use here both terms interchangeably.

A natural question that arises in the situation when some nodes are treated as latent in a Bayesian network model is how to evaluate goodness of fit, and how to parameterise such models. We will review the details of the developments and their implications in this context. Our discussion is exemplified and tested in the context of the theoretical CS&L research. Again, we stress that our considerations relating to the theory of CS&L are meant merely as an illustration of our procedure, and it is not or aim to gain extensive insight into the CS&L phenomenon.

5.1.1. Objectives

The case study in this chapter is the second one that aims at investigating the research question no. 1, namely, how marketing theories can be discovered by means of the Bayesian network approach.

In particular, this chapter has the following goals and sub-goals:

1. How can marketing theories be discovered by means of BNs? Specifically,
    a. we evaluate Bayesian networks in terms of the deductive CS&L research,
    b. we propose and evaluate new methods for:
        i. handling of latent constructs and accounting for the measurement model in BN modelling,
        ii. latent construct validation in BN modelling,
        iii. finding the dimensionality of latent constructs in BN models,

2. With regard to the added value of modelling marketing problems with Bayesian networks, we show and illustrate the potential of combination of prior knowledge with data at hand.

3. Furthermore, we pinpoint what are the strengths and weaknesses of Bayesian networks in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc

First, we evaluate Bayesian networks in the deductive CS&L research. As shown in Figure 1.3.1, developing a theory in the deductive route consists in making speculations, discussing assumptions, forming hypothetical model(s), testing them and deducing generalizations. Our speculations about the CS&L phenomenon can be seen as part of the literature overview presented in Section 3.1, therefore we focus on the remaining steps in the process: we advance possible competing hypotheses of absence or presence of direct relationships between concepts, form models and compare them with each other by means of the posterior probability measure.

Second, we propose and evaluate new methods that bring Bayesian networks closer to "good science" [Dillon *et al.*, 1997], as it is the case in second generation techniques, by enhancing the Bayesian network approach with the potentials of accounting for latent constructs and measurement models. As the first topic in this regard, we propose and evaluate a specific method for the handling of latent constructs and structural model, as well as for the accounting for the measurement model in Bayesian network modelling. More specifically, our idea of incorporating latent constructs explicitly within the measurement model consists in using a special kind of Bayesian network models, known as Naïve Bayes structures [Duda and Hurt, 1973]. To this end, we consider the use of reflective indicators. Furthermore, we show how a hidden network model can be parameterised, and evaluated in terms of its posterior probability. In order to assess the goodness of fit of the latent construct model, we apply and examine a novel method to calculate the effective dimension. Whether our approach can be deemed successful, we will judge on the basis of theoretical outcomes of the most likely model, like the nature and strengths of relationships between constructs in the structural model and by examining the relationships in the measurement models. Furthermore, we will compare our approach with the approach applied today, which is based on taking the arithmetic mean of the indicator variables and using this value as observed variable; this comparison will be based mainly on using the models as classification systems.

As the next issue in measurement modelling, we propose and examine a method of latent construct validation within the Bayesian network technology. The construct validation approach taken in this study can be seen as the extent to which an operationalization measures the concept it is supposed to measure [e.g., Cook and Campbel, 1979]. In our implementation of this definition of construct validation, we assess whether the indicator variables relate either to only one potential latent construct or to more potential different latent constructs.

Furthermore, another sub-goal of this study is to propose and evaluate a method for finding the dimensionality of latent constructs in Bayesian network models. Here, dimensionality is understood as the most likely number of states that a latent construct takes on. The assumption that underlies this objective is therefore that a concept under consideration does not involve a continuum of values, i.e., it is rather discrete ordinal variable with only several potential values. We determine the dimensionality based on its most likely measurement model.

Next, we show and illustrate the potential of combination of prior knowledge with data at hand within the Bayesian network modelling. More precisely, we consider a scenario, likely to occur in practice, in which a researcher's intention is to make use of existing theory by bringing it in the empirical validation of the model. The presumed prior knowledge in the presented example concerns values of prior conditional probabilities distributions that define relationships between a construct and its antecedents. We investigate further what is the impact of different prior knowledge on the resulting marginal likelihood of the model.

Finally, in the course of discussion, we note and collect what are the strengths and weaknesses of Bayesian networks in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc.

All the proposed methods are applied in a theoretical CS&L study set in the service industry. Since we use an existing secondary data set, we fell back on the contents of the questionnaire and operationalization of the constructs. Upon the consultation of the questionnaire and available dataset, we have decided to include four constructs in this study: Customer Satisfaction, Involvement, Trust and Loyalty.

The organisation of this chapter is as follows. In Section 5.2 we describe the collection of data, contents of the questionnaire and operationalization of constructs. In Section 5.3 we give an account of specification of assumptions and possible hypotheses. Handling of latent constructs and accounting for the measurement model in the Bayesian network framework are addresses in Section 5.4. Section 5.5 contains the discussion of the proposed construct validation procedure. In Section 5.6, we focus on determining dimensionality of latent constructs, and the results of the comparison between the competing hypothetical models are addressed in Section 5.7. Details on the implications of the most likely model in terms of the marginal probabilities for variables and strengths of relationships between constructs are the topic of Section 5.8. Our approach of handling latent constructs is compared with a standard approach in Section 5.9. We close this chapter with conclusions and implications in Section 5.10.

5.2. Data issues

5.2.1. Collection

The data being used in this study come from a telephone customer satisfaction survey among clients of a service company in Belgium (due to the legal issues, we cannot give a precise information on the name of this company and the type of service it offers). The study was aimed to investigate the extent to which the products and services offered by the company fulfil the expectations of its customers. The survey was performed via a market research company specialized in customer satisfaction and loyalty studies in October 2002. The questioning of clients ended up with a collection of 477 respondents in the dataset. Unfortunately, we do not have access to the information about the response rate, and the business profile of the respondents.

5.2.2. Questionnaire

The questionnaire contained questions of several types, including enquiries over the performance of service attributes, merchandising profile, and loyalty. For the analysis in this study we have selected constructs that on the basis of the literature overview presented in Chapter 3 can be regarded most relevant to customer loyalty, namely, customer involvement, satisfaction, trust, and loyalty.

Construct	Variable	Items
Trust	Tr1	... shows appreciation for me as customer.
	Tr2	I have trust in ...
	Tr3	... helps me always to solve possible problems.
	Tr4	... provides me with needful suggestions to use the products in the best manner.
Involvement (Inv)	Inv1	I feel involved with ...
	Inv2	I stick up for ... with my friends and the public.
	Inv3	I feel proud to be a customer of ...
	Inv4	I feel part of the success of ... in the market.
	Inv5	I share the same values as ...
Loyalty	Loy	I will remain buying from ... also in the future.
Satisfaction	Sat	How satisfied are you in general with the products and services offered by ...?

Table 5.2.1 Operationalisation of the constructs included in the study.

Two out of four variables in this study were measured with more than one item. We present the formulation of the measurement items in Table 5.2.1. Involvement and Trust were operationalized by five and four items, respectively. Trust was operationalized in terms of confidence in customer-orientation of the supplier (item Tr1), or belief that the supplier can improve the situation of the customer (Tr3 and Tr4). Items Tr1, Tr3, and Tr4 could be interpreted as related to the actual behaviour of the company, however they should rather be seen as a respondent's projection of the company's behaviour, and, consequently, as the confidence of the respondent in the relationship.[1] Measurement instrument for Involvement contained questions typical for involvement (Inv1 and Inv4), but we found also questions relating more to affective commitment (Inv3 and Inv5). Satisfaction was measured with one question related to products and services of the company in general. Similarly, Loyalty was measured in terms of the future purchase intentions also with one item.

The respondents could react how strongly they agree or disagree with the statements on the 10-point Likert-style rating scales. For Trust, Loyalty and Involvement, the value 1 stands for "completely disagree", and 10 for "completely agree". The scale for Satisfaction ranged from "very dissatisfied" to "very satisfied".

5.2.3. Data recoding

After careful inspection of the raw data, we recoded the observed variables, so as to decrease the number of values, which they take on, from ten to three. This aggregation step is advocated for the model building because the conditional probability tables should not be generally too large, unless the volume of the available data is sufficiently large for reliable parameter estimation. As follows from the example histograms in Figure 5.2.1, the responses are skewed with a low percentage of responses at the extreme values, and the majority of responses slightly above the average values. Upon careful examination of the data distribution on all the histograms, we found for almost every question that three homogenous clusters of responses could be distinguished: a cluster for responses in the range [1, 4], another cluster for responses in the range [5, 7], and another one in the range [9, 10]. More precisely, we have found that the frequency of responses for each value between 1 and 4 oscillates around 20, for responses between 5 and 7 it oscillates around 50-70, and for responses between 9 and 10 it varies between 20-40; the value 8 was by far the most frequently selected value by the respondents for almost every observed variable, and since there were relatively few responses in the range [9, 10], we have decided that the value 8 should be put together in the latter cluster. A small variance for observed variables within each cluster and high variance between clusters supports the view that the respondents regard the values inside each cluster as equivalent. As a result, the responses in the range [1, 4] received the state "low", the values in the range [5, 7] were relabelled to "moderate", and the values bigger than 7 have been renamed to "high". Correspondingly, the new labels were given to Satisfaction resulting in the states of "low satisfied" ("weakly satisfied" or "dissatisfied"), "moderately satisfied", and "highly satisfied".

[1] For instance, the question in item Tr1 should have been worded supposedly "During the last 6 months, the company showed appreciation for me as a customer" if it were to measure the actual behaviour of the company.

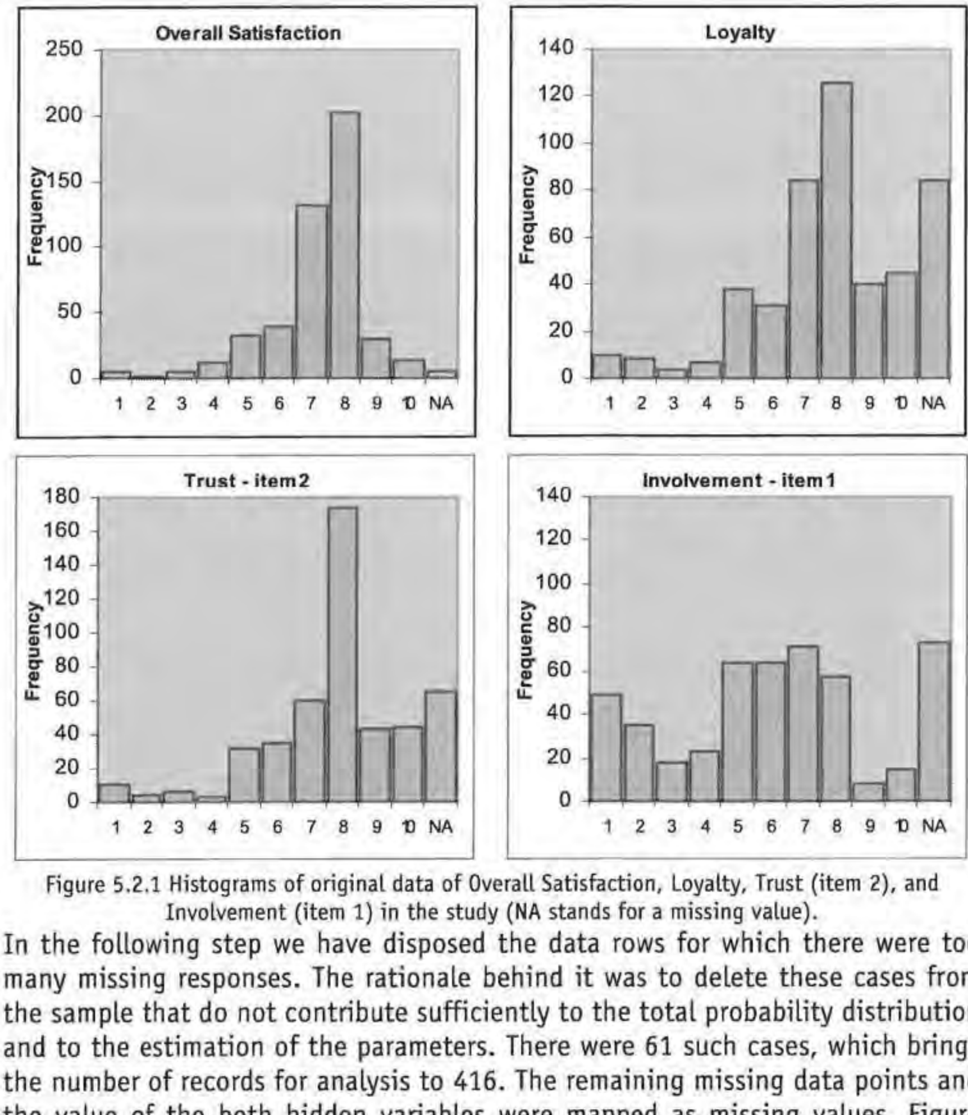


Figure 5.2.1 Histograms of original data of Overall Satisfaction, Loyalty, Trust (item 2), and Involvement (item 1) in the study (NA stands for a missing value).

In the following step we have disposed the data rows for which there were too many missing responses. The rationale behind it was to delete these cases from the sample that do not contribute sufficiently to the total probability distribution and to the estimation of the parameters. There were 61 such cases, which brings the number of records for analysis to 416. The remaining missing data points and the value of the both hidden variables were mapped as missing values. Figure 5.2.2 shows the histograms of Overall Satisfaction and Loyalty after the aggregation step.

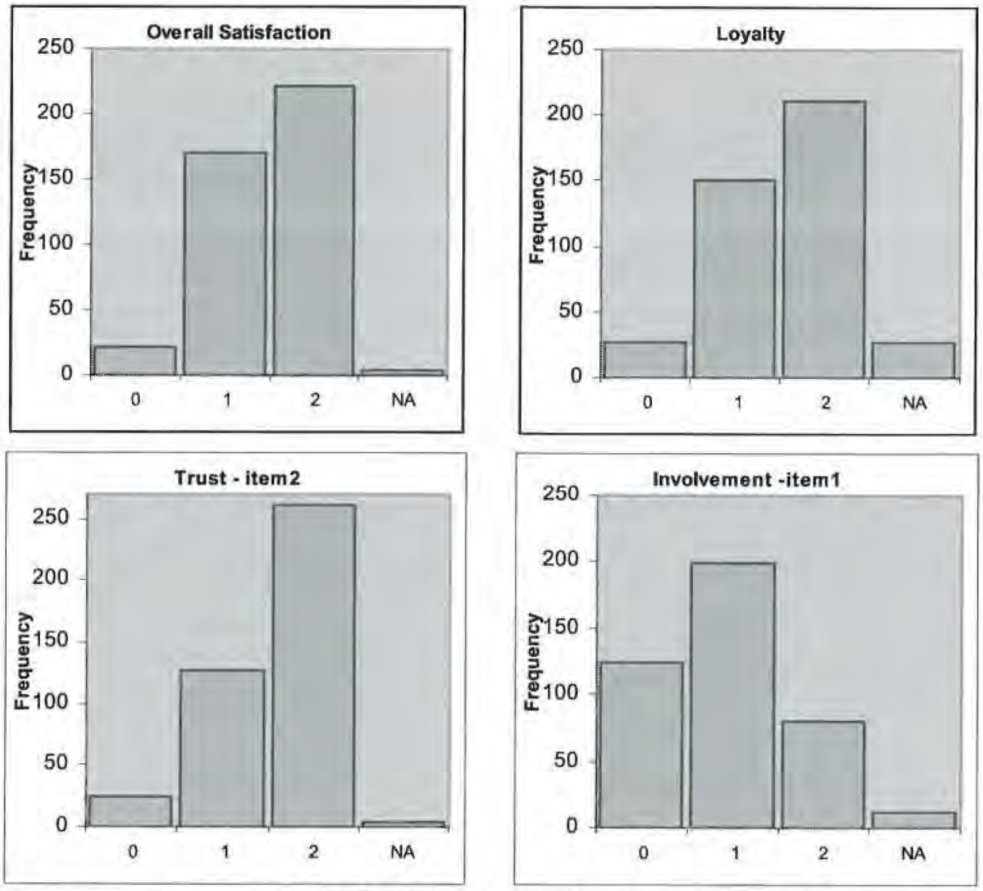


Figure 5.2.2. Histograms of Overall Satisfaction, Loyalty, Trust (item 2), and Involvement (item 1) after aggregation (NA stands for a missing value).

In the same way we have also aggregated all the manifest variables relating to Trust and Involvement.

5.3. Specification of alternative models

In this section we present several hypothetical models of Customer Loyalty that will serve as the background for our discussion and application of Bayesian networks in latent construct modelling. We would like to stress that these models are oversimplified and far from being complete representation of the Customer Loyalty phenomenon. Nevertheless, in our opinion, the five models that we present can be perceived as *a priori* most likely theoretical models of relationships between Customer Satisfaction, Involvement, Trust and Loyalty, i.e., the four constructs present in our data set. The specification of these models is to a great extent based on a review of previous research in the CS&L literature and on expertise of the market research company that performed the data collection, rather than on our own subjective knowledge. However, whenever appropriate, we make also references to the selected relevant previous studies that we found in the CS&L literature, but refer to Section 3.2 for a fuller account of this relevant literature.

We specify the following models by presenting our propositions concerning presence or absence of specific direct relationships between a pair of involved constructs. Collectively, these propositions form a hypothetical theoretical model of CS&L. We stress that we refer to these relationships as propositions, and not as hypotheses, since we agree that formulating a scientific hypothesis requires more evidence and support in the accumulated body of knowledge than formulating a proposition [Sekaran, 1992].

In all the considered models we have assumed that customer loyalty is the ultimate dependent concept in the domain, because of its value in as a proxy for profitability [Reichheld and Sasser, 1990; Fornell *et al.*, 1996], therefore we test it only as a child of the remaining constructs.

Let us start with the model presented in Figure 5.3.1a. Firstly, it postulates that Loyalty is determined directly by Involvement.

Morgan and Hunt [1994] demonstrate a negative relationship between Trust and propensity to leave. Anderson and Weitz [1989] have found evidence that Trust is key to maintaining continuity in conventional channel relationships. Similarly, Doney and Cannon [1997] found that Trust of the supplier firm and of the salesperson increase a buyer's anticipated future interaction with the supplier.

There is a debate concerning the causal ordering between Satisfaction and Trust [Geyskens *et al.*, 1998]. We assume that Satisfaction is an antecedent of Trust. Therefore Loyalty is a consequence of Satisfaction but indirectly through Involvement and Trust. Also, one of the building blocks of engagement is Trust [Smith and Rutigliano, 2003], hence the link from Trust to Involvement, which is conceptually close to Engagement. Model 1, which encodes these relationships graphically, is postulated by the research agency that collected the data.

However, it is likely that no direct link exists between Involvement and Trust. Therefore, in Model 2 shown in Fig. 5.3.1b we propose the same qualitative interdependencies as in Model 1, with the exception that neither Trust influences Involvement or the other way around.

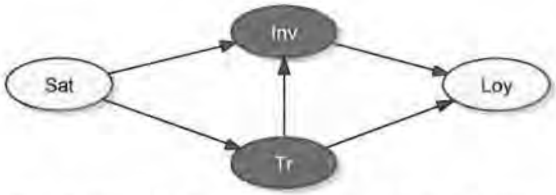
Model 3 is another simplification of Model 1, in which there is no link between Trust and Involvement and, additionally also no link between Satisfaction and Involvement. This model is presented graphically in Figure 5.3.1c.

The next hypothetical model, Model 4, assumes that Involvement has a direct influence on Loyalty as well as on Satisfaction. In this model Satisfaction is also antecedent of Loyalty, but Trust mediates this link. This model is shown in Figure 5.3.1d.

Finally, the last hypothetical model shown in Figure 5.3.1e is Model 5. The model is an alternative to Model 1, and is different by the direction of the influence between Involvement and Trust. This structural model is based on the finding that Involvement leads to Trust, rather than the other way around [Teichert and Rost, 2003].

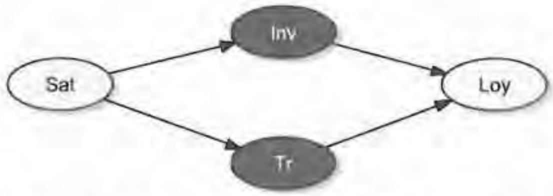
In Figure 5.3.1, besides the graphical representation of the discussed models, we show the formula for the joint probability distribution that each model entails. We can view these models as structural, or latent variable, models, since the measurements are not shown.

a) Model 1



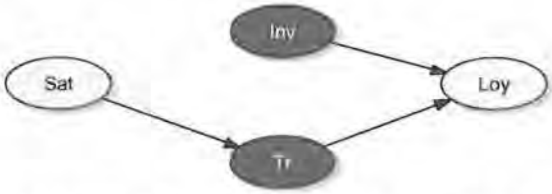
$$p(Sat, Tr, Inv, Loy) = p(Sat)\, p(Tr \mid Sat)\, p(Inv \mid Sat, Tr)\, p(Loy \mid Inv, Tr)$$

b) Model 2



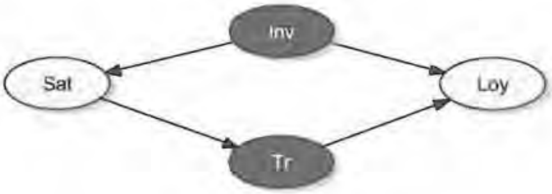
$$p(Sat, Tr, Inv, Loy) = p(Sat)\, p(Tr \mid Sat)\, p(Inv \mid Sat)\, p(Loy \mid Inv, Tr)$$

c) Model 3



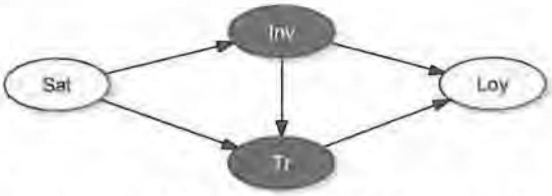
$$p(Sat, Tr, Inv, Loy) = p(Sat)\, p(Tr \mid Sat)\, p(Inv)\, p(Loy \mid Inv, Tr)$$

d) Model 4



$$p(Sat, Tr, Inv, Loy) = p(Inv)\, p(Sat \mid Inv)\, p(Tr \mid Sat)\, p(Loy \mid Inv, Tr)$$

e) Model 5



$$p(Sat, Tr, Inv, Loy) = p(Sat)\, p(Inv \mid Sat)\, p(Tr \mid Sat, Inv)\, p(Loy \mid Inv, Tr)$$

Figure 5.3.1 Alternative structural models considered in the study.

5.4. Latent constructs modelling

In this section, we will present our methods of accounting for latent constructs and measurement model within the Bayesian network modelling. Let us first discuss our proposed approach to measurement modelling, followed by topics that are critical in our methodology, i.e., the EM algorithm and the effective dimension of the hidden network model.

5.4.1. Measurement models

Measurement is the process by which a theoretical concept is linked to one or more latent variables, and these variables are linked to observed variables [Bollen, 1989]. There are generally four steps in the measurement process: 1) developing a theoretical definition, 2) identifying the dimensions and latent variables to represent it, 3) developing an operational definition, 4) specifying the relation between the measures and the latent variables. We will focus on the last step.

A measurement model specifies a structural model connecting latent variables to one or more measures of observed variables [Bollen, 1989]. So, the measurement model describes the relation between the measure and the latent construct. In the theory of construct measuring, the relation between latent constructs and indicators is referred to by the discussion of correspondence rules [see Bagozzi, 1999, p.321]. It is important to incorporate the latent constructs as latent variables in the model, since it brings us closer to "good science" [Dillon *et al.*, 1997].

The case with hidden variable Bayesian network models involves more complexity than models with missing data. The hidden variable models can be compared to missing data models in which some variables are never observed, i.e., all cases are treated as missing, whereas in missing data models only some data points are not reported for variables.

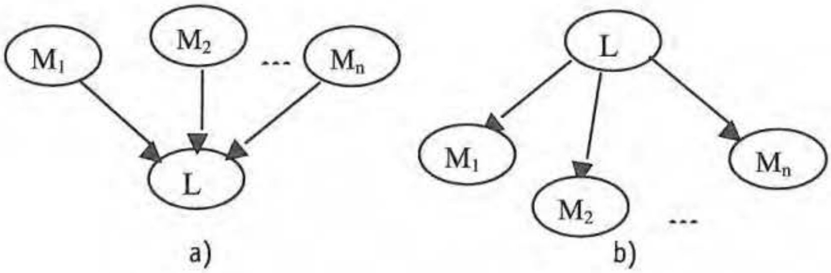


Figure 5.4.1 Examples of a) cause indicators, and b) effect indicators, where *L* is a latent construct and *M* are indicator variables.

We have assumed that indicator variables depend on a latent variable. In this sense they can be called *effect* indicators (Fig. 5.4.1b), as opposed to *cause* indicators (Fig. 5.4.1a) [Bollen, 1989, p. 65]. The effect indicators, also called reflective indicators, are assumed to be caused by a latent variable, whereas cause indicators, known also as formative indicators, are assumed to cause a latent variable. Bollen [1989] notes that most researchers in social sciences assume that indicator variables are effect indicators, despite many instances in which cause indicators are appropriate.

Deciding whether an indicator is an effect or a cause indicator can be troublesome. Often, it can be determined by establishing a causal or temporal priority between an indicator and a latent variable. We assume here that the indicators are effect indicators.

So, in our treatment of measurement models we shall assume that the indicator variables are mutually independent given the value of the latent variable. They are however still interdependent marginally when the value of the latent construct is not known. Of course, in principle, the value of the latent variable is actually never known, however once the model is built and estimated we can assume the value of latent variables to test various hypotheses and perform probabilistic simulations just as if we were able to observe the state of the latent construct. The assumed dependency model can be represented with a graph shown in Figure 5.4.2. This form of probabilistic interdependencies is known as Naïve, or simple, Bayes model [Duda and Hart, 1973].

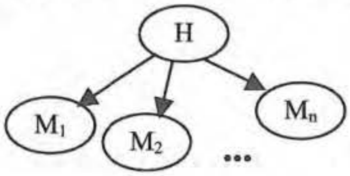


Figure 5.4.2. The Naive Bayes conceptualization of measurement of the latent variables.

The joint probability distribution over the latent variable and indicator variables can be according to the Naïve Bayes model given as:

$$p(H, M_1, ..., M_n) = p(H)\prod_{i=1}^{n} p(M_i \mid H), \qquad (5.1)$$

where H is the latent variable, and M_i are the effect indicator variables that measure the latent variable H.

The parameterisation of the measurement model requires that the conditional probabilities $p(M_i|H)$ as well as the unconditional probabilities $p(H)$ in the formula (5.1) are estimated. Clearly, this parameterisation would be trivial if we were in possession of observed data for H. In that case these probabilities could have been estimated as relative observed frequencies of M_i given the value of H. Unfortunately, we have no data on H and therefore we can only extrapolate these probabilities by means of optimisation techniques.

The most frequent optimisation techniques used in statistics to deal with missing data make use of the formulation of this problem as finding the maximum of the likelihood function. Let $p(\mathrm{x}|\Theta, B_s)$ be a density function, where $\Theta = \cup_{i,j,k}\theta_{ijk}$ is the set of all the parameters in the Bayesian network structure. Also, let $D=\{\mathrm{x}_1, \ldots, \mathrm{x}_N\}$ be the observed data set of N tuples (cases) where each x_l is a tuple of the form $(x_1,\ldots,x_n)$ that assigns values to the variables $\mathrm{X}=\{X_1, \ldots, X_n\}$. Then, assuming that all the cases in D are independent and identically distributed given the model, the likelihood function $L(\Theta|D, B_s)$ can be expressed formally as

$$p(D \mid \Theta, B_s) = \prod_{l=1}^{N} p(\mathrm{x}_l \mid D, B_s) = L(\Theta \mid D, B_s)\,, \qquad (5.2)$$

In other words, the likelihood function is the probability of observing the data D given a current parameterisation Θ and a Bayesian network structure B_s. So, if we assume that the Bayesian network structure B_s is fixed, the goal of the optimisation is to find such parameter values $?_{ijk}$ that maximize the function (5.2).

Usually, one computes the natural logarithm of the function in (5.2) to avoid the troubles with the numeric precision of the calculation (with big datasets the likelihood value can be extremely small), so we arrive at the loglikelihood function

$$LL(\Theta \mid D, B_s) = \log \prod_{l=1}^{N} p(\mathrm{x}_l \mid \Theta, B_s) = \sum_{l=1}^{N} \log p(\mathrm{x}_l \mid \Theta, B_s)\,, \qquad (5.3)$$

The loglikelihood has also a statistical interpretation that the higher the likelihood of the model, the closer this model is to modelling the probability distribution in the data D.

To this end, one can apply any optimisation technique, the most popular of which include the gradient-based techniques, the Monte Carlo sampling techniques (e.g., Gibbs sampling), or the expectation-maximization (EM) algorithm [Dempster *et al.*, 1977]. Lauritzen [1995] adapted the EM algorithm for the special case of Bayesian networks. We will in the experiments in the remaining part of this chapter use our own implementation of the EM algorithm along with our own implementation of the metrics.

5.4.2. EM algorithm

We have chosen the EM optimisation because it is quite simple, both theoretically and computationally, and it is known to converge relatively fast to the local optimum.

In general, the goal of the EM algorithm is to find parameter values θ that maximize the likelihood function $p(D \mid \Theta)$, where D is some given set of data. Applied in the Bayesian network setting, the goal of the EM learning is to find such values of the conditional probabilities Θ so as to maximize the likelihood of the model $p(D \mid \Theta, B_s)$ given the data and a Bayesian network structure B_s.

Let $\widehat{\Theta}$ denote the instantiation of parameter values such that maximize the likelihood function $p(D \mid ?, B_s)$:

$$\widehat{\Theta} = \arg\max_{\Theta} p(D \mid \Theta, B_s), \tag{5.4}$$

The assignment $\widehat{\Theta}$ is called the *maximum likelihood* (ML) estimates of the true parameters Θ.

The idea of the EM learning is to find the expected values $E(N_{ijk})$ by performing inference in the network for each missing variable in every case, and use these expected values as if they were the observed values to compute the probability as . The algorithm starts with an initial random parameterisation of θ_{ijk}, let's call it θ'_{ijk}. Next, we compute the expected sufficient statistics for a complete data set, where expectation is taken with respect to the joint distribution for X conditioned on the assigned configuration.

$$E_{p(\mathrm{x} \mid D, \theta_s, B_s)}(N_{ijk}) = \sum_{l=1}^{N} p(x_i^k, \pi_i^j \mid y_l, \theta_s, B_s), \tag{5.5}$$

where y_l is the possibly incomplete lth case in D. When X_i and all the variables in the X_i's parent set π_i are observed in case x_l, the term $p(x_i^k, \pi_i^j \mid y_l, \Theta_s, B_s)$ for this case requires a trivial computation: it is either zero or one. Clearly, it is one when X_i is observed in state k and the parents are in configuration j; and zero when X_i is observed in state other than k, or in state k but the parents are in the configuration other than j. Otherwise, i.e., if either the value of X_i is missing, or at least one of the parents are missing in the data for the case x_l, then we can use any Bayesian network inference algorithm to evaluate the term. This computation is called the *expectation* step, or the E-step, of the EM algorithm.

In the next step we are using the expected sufficient statistics $E_{p(\mathrm{x} \mid D, \theta_s, B_s)}(N_{ijk})$ received in the E-step just as if they were the actual sufficient statistics counted from a complete random dataset D_c. The calculation to be performed in this step depends on whether we are doing the ML ("maximum likelihood") or the MAP ("maximum a posteriori") parameter estimation. In the ML configuration can be reached when we do not use any prior estimates of the parameters and therefore they are calculated as

$$\theta_{ijk} = \frac{E_{p(x|D,\theta_s,B_s)}(N_{ijk})}{\sum_{k=1}^{r_i} E_{p(x|D,\theta_s,B_s)}(N_{ijk})} = \frac{N'_{ijk}}{N'_{ij}}, \tag{5.6}$$

The MAP configuration can be computed as

$$\theta_{ijk} = \frac{\alpha_{ijk} + E_{p(x|D,\theta_s,B_s)}(N_{ijk})}{\sum_{k=1}^{r_i} (\alpha_{ijk} + E_{p(x|D,\theta_s,B_s)}(N_{ijk}))}, \tag{5.7}$$

The ML estimate of the probabilities θ_{ijk} can be therefore seen as a special case of the MAP estimation, in which we are *a priori* completely uninformed and rely only on the data and let the priors be $\alpha_{ijk} = 0$. The assignments above constitute the maximization step of the EM algorithm. At the end of each run, the likelihood function is guaranteed to be higher than in the previous run. The algorithm repeats the calculations in the E and M steps until the relative difference in the loglikelihoods $LL(D|\theta, B_s)$ in the two consecutive iterations changes only to a small extent.

When the difference between the loglikelihoods $LL(D|\theta, B_s)$ of the two consecutive runs becomes small enough, we can expect that the algorithm virtually reached a local optimum, so further iterating would improve the likelihood and probabilistic parameters only minimally. Whether the optimum reached is also the global one, assuming there is one global optimum, depends on the initial parameter values, therefore the whole procedure, including the E and M steps, should be repeated a sufficient number of times with possibly maximally varied parameter values in order to maximize the chance of localising the global optimum.

It is known that the algorithm usually slows down in finding the optimum close to the optimal values. The consecutive iterations do not improve then the value of the likelihood function remarkably any more. One of the methods to speed up the convergence is to switch to the gradient-based algorithm when the EM algorithm slows down [Thiesson, 1995].

When working with small databases it may also happen that some combinations of parents' states for a variable X_i can never be observed. Then, the nominator as well as the denominator in Expression 5.7 becomes zero. For that reason, in practice, to prevent from the division-by-zero error and to avoid zeros in the conditional probability tables, one assumes the value of a_{ijk} equal one.

5.4.3. Effective dimension of the model

Recall from the section 3.6.1 that the marginal loglikelihood of a model in case of hidden variable models can be estimated by the approximation called the Bayesian Information Criterion (BIC) as well as by the Cheeseman-Stutz (CS) approximation [Chickering and Heckerman, 1997]. In a number of simulation studies, these two measures have turned out to be the best approximations [Chickering and Heckerman, 1997]. Both the BIC and CS approximations of the marginal likelihood are based on large-sample properties of the probability distribution. To avoid confusion, it is useful to mention that both these measures already account for the complexity of the model.

The computation of the BIC and CS scores requires thus calculation of the model's dimension. In the case of models with missing data, the dimension is equivalent to the structural dimension d, i.e, it can be computed as $d = \sum_{i=1}^{n} q_i(r_i - 1)$, where q_i is the number of different combinations of parents' values for node i, and r_i is the number of states for node i. For example, when we have a model $X \rightarrow Y$, with binary random variables X and Y, then the structural dimension is d=3. However, when dealing with hidden variable models, the correct approach involves calculation of the effective dimension d' instead of the structural dimension [Geiger *et al.*, 1996]. This is because some of the structural parameters are redundant. For instance, in a network $H \rightarrow X$ where both H and X are binary, X is observed and H is hidden, there is only one non-redundant parameter, whereas the number of parameters that refer to the structural dimension amounts to three. The value of the effective dimension can be lower than the usual number of the parameters.

Now, the BIC approximation of the marginal loglikelihood of the hidden variable model B_s can be expressed as follows:

$$\log p(D \mid B_s) \approx \log p(D \mid \tilde{\Phi}, B_s) - 0.5 d' \log N, \tag{5.8}$$

where $p(D \mid \tilde{\Phi}, B_s)$ is the likelihood of the model in the ML configuration of the model's parameters $\tilde{\Phi}$, N is the number of observations, and d' is the effective dimension of the model.

Corrected for a model with hidden variables, the CS approximation is given by

$$\begin{aligned} \log p(D \mid B_s) \approx \log p(D' \mid B_s) - \log p(D' \mid \tilde{\Phi}_{B_s}, B_s) + \\ \log p(D \mid \tilde{\Phi}_{B_s}, B_s) + 0.5(d' - d) \log N \end{aligned} \tag{5.9}$$

where d is the structural dimension, and d' is the effective dimension.

Calculation of the effective dimension is computationally not easy. In fact, it is an NP-hard problem [Settimi and Smith, 1998]. It involves computation on matrices, whose size grows exponentially with the number of nodes in a hidden network model. Settimi and Smith [1998] noticed that the calculation of the effective dimension of the hidden node model can be split up into a sum of effective dimensions around the Markov neighbourhood of each hidden node separately. This can in some cases, dependent on the network structure, reduce the complexity of the task substantially, but in general the task is complex and cannot be solved in a linear time. To calculate the effective dimension, we applied an S-PLUS procedure described in [Rusakov and Geiger, 2003] that implements the observation of Settimi and Smith [1998]. We note, that the calculation of the effective dimension is at present a serious bottleneck in state-of-the-art applications of Bayesian networks. Some authors use the structural dimension instead, when precise determination of the MLL approximations is not critical for the research objective [see e.g., Friedman, 1998].

5.5. Construct validity

Before a multi-item scale can be used it should be evaluated for accuracy and applicability [Greenleaf, 1992]. This involves an assessment of reliability, validity, and generalizability of the scale. We will assume in this study that the scales used are reliable and generalizable, since it is not so important in the whole discussion, and so we will focus primarily on the validation of the scale.

Validity can be defined as the extent to which differences in observed scale scores reflect true differences among objects on the characteristic being measured, rather than systematic or random error [Malhotra, 1993]. Validity can be assessed by examining content validity, criterion validity and construct validity [Singh, 1991]. Among the three types of validity, we propose a way to examine the construct validity. The last type of validity refers to the construct validity, which attempts to address the question of what construct or characteristic the scale is actually measuring [Malhotra, 1993]. It can be defined as the extent to which an operationalization measures the concept it is supposed to measure [e.g., Cook and Campbel, 1979]. When assessing construct validity the researcher needs to answer why the scale works and what deductions can be made concerning the underlying theory. Construct validity includes convergent, discriminant, and nomological validity. Convergent validity is the extent to which the scale positively correlates with other measures of the same construct. Discriminant validity is the extent to which a measure does not correlate with constructs with which it is supposed to differ. Lastly, nomological validity is the extent to which the scale correlates in theoretically predicted ways with measures of different but related constructs.

We will conduct a procedure of construct validation using also the Bayesian network approach. Our method is inspired by the idea of detecting hidden variables in Bayesian networks with the Bayesian scoring metric [Heckerman *et al.*, 1997; Friedman, 1998; Elidan *et al.*, 2000]. The method that we propose here is conceptually close to the Latent Class Factor Analysis (LCFA) of Vermunt and Magidson [2004], however it differs in some respects, e.g., we do not put any regression-type like constraints on the item conditional probabilities. Some other similar approaches have been earlier proposed by Goodman [1974] and Hagenaars [1988]. In our approach, we verify whether the indicator variables that were aimed to measure a specific latent construct do measure in fact one and the same conceptual construct. In order to test it, an experiment was designed in which we hypothesized the existence of two hypothetical latent constructs: one of them was focal latent construct (Trust or Involvement), and the other one was an unknown, arbitrary latent construct. Next, we scored various models that were different from each other in the links between latent constructs and indicator variables. Here again, we use approximation of the marginal likelihood to score each potential model and to compare them. The models with the highest marginal likelihood are preferred over the models that score lower. Now, our idea is that if in the most likely model indicators are related with only one latent construct, then it indicates that these indicators indeed measure one and the same concept; otherwise, i.e., if a model, in which some indicators are "caused" by the second latent construct, is more likely, then it suggest that the indicators are measures of two different constructs. The approach here can be considered two-stage estimation as opposed to one-stage estimation in that we first attempt to assess the quality of the measurement items, and next we estimate the quality of the entire model with a subset of the items [Fornell and Yi, 1992]. Hence, we consider a setting that in the literature is referred to as pure, or unidimensional, measurement model [Anderson and Gerbing, 1982, 1988], because each indicator is a direct effect of exactly one latent variable.

In these experiments, we have a priori assumed that both Trust, and Involvement, as well as the potential second hidden variable are ternary. The results are contained in Table 5.5.1. The priors α_{ijk} were selected with the equivalent sample size of 5, and where the probabilities $p(x_i \mid \mathrm{pa}_i, B_s)$ were uniform. This means that the conditional distributions $p(x_i, \mathrm{pa}_i|B_s)$ for each node i, and each combination of states of its parents are uniform, and the equivalent sample size of 5 means that these uniform probabilities were based on the hypothetical, imaginary size of prior sample of size 5 cases. This is a way to express that our prior beliefs on these distributions are supported by some knowledge. We would like to notice that the size of 5 is taken rather arbitrarily. The approximations shown in the table are the maximum values obtained from among 100 runs of the EM algorithm, each with different random initial parameters. We show also the effective (*de*) and the structural (*ds*) dimension of the models, and the rankings of each model when compared by the CS (rcs), and the BIC (rbic) approximations in the rightmost columns.

All the remaining models that are not shown in this table are symmetrical and equivalent to those shown in this table, so they do not need to be evaluated. We note that the absolute value of the approximation of the marginal likelihood is not of most significance. Here, only the marginal likelihood value of one model relative to other models is taken into account. The maximum values of both scores are typed in bold.

As we can see in Table 5.5.1, the differences in the CS and BIC scores for a specific model configuration can result from the limited sample size, as both the scores approximate the marginal likelihood function in the case of large samples. Furthermore, the structural dimension (ds) equals 28 for all models. Also, for each hypothetical model this dimension is different from the effective dimension (de), which varies from 16 to 26. This result shows that we should calculate the value of the effective dimension and use this value, since not doing so can potentially bias the value of the marginal likelihood approximation, and alter the ranking of the alternative models.

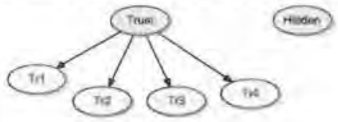
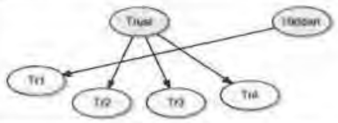
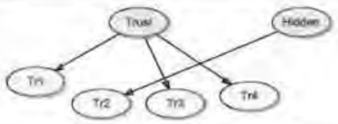
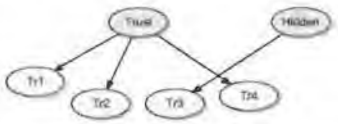
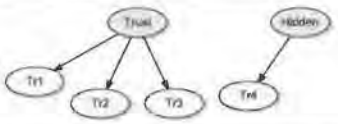
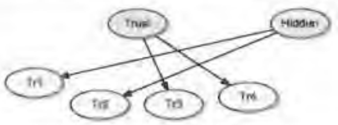
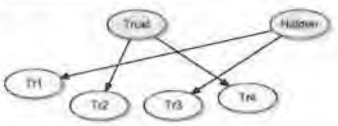
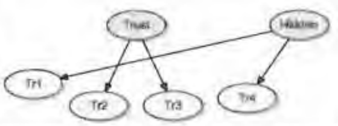
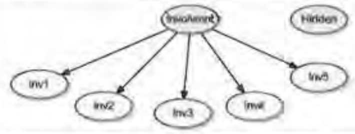
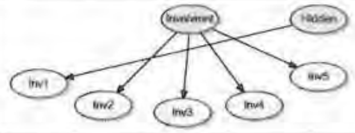
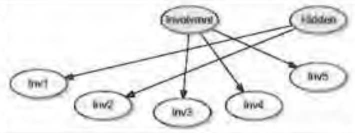
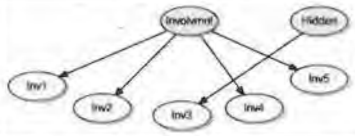
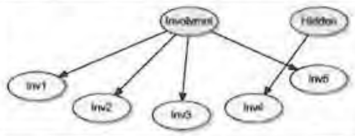
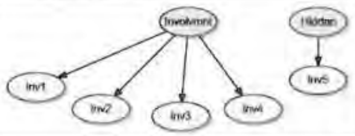
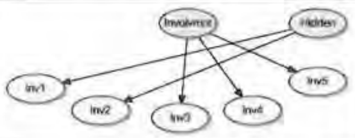
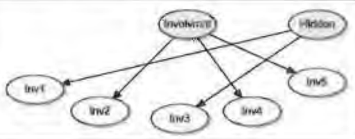
No.	Model	CS	BIC	de	ds	rcs	rbic
1.		**-1353.165**	**-1269.174**	26	28	1	1
2.		-1428.32	-1353.965	22	28	8	5
3.		-1404.206	-1353.655	22	28	5	6
4.		-1421.893	-1350.222	22	28	6	4
5.		-1374.720	-1333.948	22	28	2	2
6.		-1382.443	-1339.981	16	28	3	3
7.		-1396.229	-1356.022	16	28	4	7
8.		-1421.180	-1362.942	16	28	7	8

Table 5.5.1. Various measurement models for Trust (Note: *CS* and *BIC* stand for the Cheeseman-Stutz and the Bayesian Information Criterion approximation of the marginal loglikelihood, respectively; *de* is the effective dimension, *ds* is the structural dimension of the models).

Most importantly, we can observe that the most likely measurement model, both in terms of the CS as the BIC scoring function, among the considered models is the model no. 1. This model represents the assertion that the four observed variables are all effect indicators of one latent construct. Hence, it is most likely that there is only one hidden node, which is the parent of *all* the indicator variables. We can also conclude that given the face validity of the indicators this latent construct should be indeed, in our opinion, regarded as Trust. This is a very important finding for our research.

No	Model	CS	BIC	de	ds	rcs	rbic
1.		**-1513.608**	**-1471.370**	32	34	1	1
2.		-1677.784	-1600.388	28	34	3	2
3.		-1681.493	-1622.701	28	34	4	4
4.		-1741.129	-1679.484	28	34	12	11
5.		-1702.203	-1653.241	28	34	6	7
6.		-1651.345	-1601.140	28	34	2	3
7.		-1700.221	-1649.167	28	34	5	5
8.		-1745.342	-1703.326	28	34	13	14

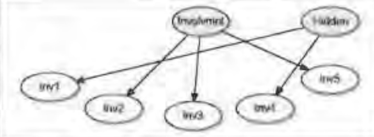
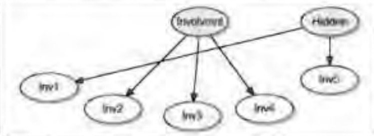
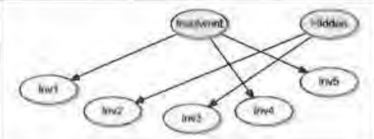
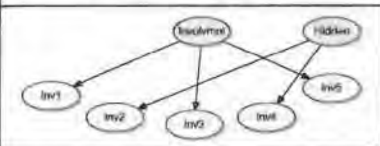
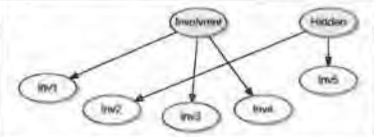
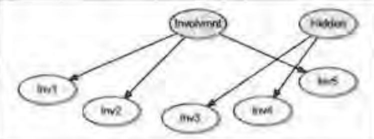
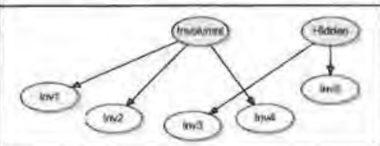
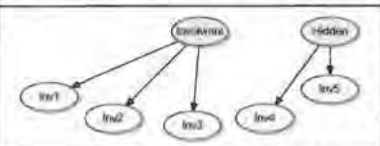
9.		-1747.667	-1693.289	28	34	14	13
10.		-1704.469	-1652.962	28	34	7	6
11.		-1736.208	-1686.829	28	34	11	12
12.		-1762.619	-1710.181	28	34	16	16
13.		-1715.977	-1676.421	28	34	9	10
14.		-1706.854	-1673.061	28	34	8	9
15.		-1750.667	-1705.828	28	34	15	15
16.		-1717.073	-1670.271	28	34	10	8

Table 5.5.2. Possible measurement models for Involvement.

Items Tr1, Tr2, and Tr3 are together the second most likely operationalization of one latent construct, so they must be related to each other. Further analysis of other likely models can provide interesting insight into the issue of potential multidimensionality of Trust. For instance, we can see that among models 6-8, in which pairs of indicators are considered as measurements of two different constructs, model no. 6 is most likely (it is also the third most likely model of all models); this can indicate that items Tr1 and Tr2 are measurements of one Trust dimension, while Tr3 and Tr4 are measurements of another dimension of Trust.

While it is difficult to interpret the first dimension, the second one could be viewed as trust in help and benevolence of the product provider.

A similar procedure can be applied to test the operational validity of the indicators of Involvement. We consider 16 unique models with five effect indicator variables. The results presented in Table 5.5.2 suggest that the most likely potential measurement model of Involvement is the model 1 encoding that all the five indicator variables can be commonly the effect indicators of one latent construct. This means that the five items in Table 5.2.1, that we identified as potentially relating to different constructs, such as affective commitment, organizational citizenship behaviour and involvement, are probably just different aspects of one concept that most likely is Involvement.

Among the four-item instruments, models with indicators Inv1-Inv4, and Inv2-Inv5 are most likely. Besides, among three- and two-item scales, the model in which items Inv1-2 relate to one concept, and items Inv3-5 relate to another one, and the model in which items Inv1 and Inv5 relate to one concept, and items Inv2-4 relate to another concept are very likely.

Let us summarize findings from Tables 5.5.1-2. We can notice that for both Trust and Involvement, it is most likely that all the indicators included in the measurement instrument are indicators of one latent construct. This would suggest that models in which more indicators are attached to one latent construct would always score higher. However, we can see in both tables that this conjecture is not true, as for Trust some two-item measurement models (models 6 and 7) are more likely than three-item instruments (models 2-4, taking the CS approximation as superior); similarly, as regards Involvement, some three-item instruments (e.g., models 7, 10, 13, 14, 16) are more likely than four-item scales (i.e., model 4).

We observe that there exist differences between the BIC and CS approximations of the marginal likelihood of considered measurement models, causing inconsistencies in the ranking, but these differences are small and are responsible for discrepancy only in terms of at most three positions.

In order to evaluate the results of the presented method of construct validation in an objective way, we should compare them with the standard procedure applied to test unidimensionality and reliability of latent constructs in CS&L research. Because this analysis can be done only for fully observed data, we have discarded the cases with missing values using the listwise deletion; as a result the number of cases used in this analysis is 271. Furthermore, we have used the original, not aggregated data in the analysis.

We have first performed factor analysis on the scales for Trust using the principal components extraction method. The results are shown in Table 5.5.3. The Kaiser-Meyer-Olkin (KMO) measure of the sampling adequacy was 0.822 and could be seen as satisfactory for factor analysis to proceed. The Bartlett's test of sphericity was significant, which means that the correlation matrix was different from the identity matrix. Based on the value of the total variance explained by the models with different numbers of potential factors, we found that models with the number of factors greater than 1 do not contribute significantly more to the total variance explained. That means that the items manifest similar characteristics and can be grouped together as relating collectively to one factor. Total variance explained for the one-factor solution for the Trust items was 72.057%. Factor loadings can be read from Table 5.3.3.

Construct & scale	KMO adequacy	Total variance explained	Factor loadings*	Reliability (Cronbach's α)
Trust	0.822	72.057%		0.8702
Tr1			0.868	
Tr2			0.859	
Tr3			0.862	
Tr4			0.805	
Involvement	0.902	80.91%		0.9412
Inv1			0.871	
Inv2			0.888	
Inv3			0.923	
Inv4			0.919	
Inv5			0.897	

Table 5.5.3 Results of the factor and reliabilty analysis (* only one factor per construct was extracted).

Next, we have performed the same analysis for the scale of Involvement. The KMO measure of sampling adequacy amounted to 0.902. For the scale items of Involvement we have also found that one-factor solution explains as much as 80.91% of total variance, with more factors improving the total variance explained only to a small extent. We can therefore conclude that, based on the measurement scales used, both Trust and Involvement are unidimensional constructs.

Once we established that both scales refer to two different factors, we were able to assess their reliability in terms of Cronbach's alphas. Cronbach's alpha measures how well a set of items (or variables) measures a single unidimensional latent construct. If data have a multidimensional structure, Cronbach's alpha will usually be low. The results of this analysis are contained in Table 5.3.3. The Cronbach's alpha for the scale of Trust amounts to 0.8702, and for Involvement it is 0.9412. These values are very high, and since in most social science applications a reliability coefficient of .80 or higher is considered as acceptable, we should regard the scales as reliable measurements of the two latent constructs.

Consequently, the results of the factor analysis, followed by the analysis of reliability of both scales are consistent with our proposed method of construct validation in the Bayesian network framework. Both in the classical approach and the Bayesian network approach, the 4-item scale used to capture Trust and the 5-item scale used to measure Involvement are reliable and show that the items refer to one construct, or factor.

It is important to note here also that a similar procedure can be used in more complex models to probe for the existence of other potential latent constructs in a theory in focus. In that case, if the network structure augmented by the introduction of a latent construct (of course without their equivalent indicator variables) would represent higher value of the likelihood, then this might be an indication that this new, previously not considered construct, can potentially play an important role in the theory under consideration. By looking at relationships between this new construct and the remaining constructs, and the CPT's, we can also get an idea what omitted concept the construct should represent [see e.g., Heckerman *et al.*, 1999]. This capability presents itself as another advantage over SEM modelling, since this capability is not present in SEM modelling.

5.6. Dimensionality of latent constructs

We have concluded the previous section by proposing that the developed construct validation approach can be suitable for validation of multidimensional nature of latent constructs. In order to avoid confusion, we define dimensionality as a number of states that a latent construct can take.

In order to be able to compute the marginal likelihood in the previous section, we had to assume a fixed number of values that latent construct can take. We assumed that this number amounts to three. Here, we will release this restriction and try to find the best number of states of a latent construct. A similar approach, but with a different objective, was taken by Cheeseman and Stutz [1995], who used the Naïve Bayes models to carry out unsupervised learning of hidden clusters in the multidimensional data.

Formally, the problem of finding the dimensionality of a latent construct H can be formulated as finding the number of states k that maximizes the marginal loglikelihood approximation function:

$$k = \arg\max_{\lambda} p(D \mid B_{s\lambda}), \qquad (5.10)$$

where $\lambda = 2, \ldots, n$ is a specific value of states of H for which we evaluate the approximation function. Here, again the marginal likelihood $p(D|B_s)$ can be approximated by a scoring function that could be either the CS function or the BIC function.

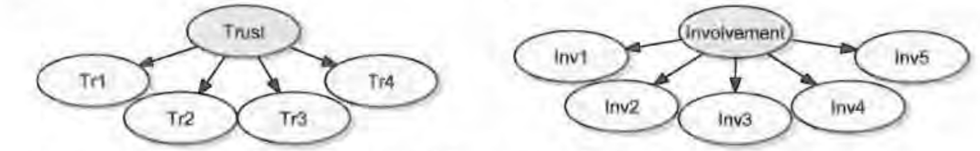


Figure 5.6.1 Structure of the model for a) Trust, and b) Involvement. The number of states for latent constructs (in grey) varies in the experiment from 2 to 6.

In the following tables we show the BIC and CS scores for the measurement models of constructs Trust and Involvement found as most likely in the previous section with varying number of states λ. The structures of these models are shown in Figure 5.6.1.

k	CS	BIC	de
2	-1316.235	-1294.178	17
3	**-1292.000**	**-1269.392**	26
4	-1312.997	-1289.489	35
5	-1325.932	-1312.159	44
6	-1342.981	-1334.805	53

Table 5.6.1 Different dimensionalities of Trust.

For Trust, as it appears from Table 5.6.1, the most likely number of states is reached when λ=3, because both the CS score and the BIC score are maximum for λ=3. We can therefore conclude that the dimensionality of Trust is 3, and that it can be best represented with 3 states that can hypothetically receive the interpretations "low", "moderate" and "high".

K	CS	BIC	de
2	-1734.6364	-1664.3499	21
3	**-1564.0738**	**-1476.0183**	32
4	-1586.2430	-1496.6237	43
5	-1609.6065	-1519.6866	54
6	-1631.8814	-1544.4064	65

Table 5.6.2. Different dimensionalities of Involvement.

As regards Involvement, we can conclude that given the Naive Bayes measurement model the dimensionality of this construct equals also 3, probably with states representing the same levels as for Trust.

As a potential critique on the presented approach, we should mention that models that postulate three states of latent constructs could be preferred over models having other number of states than three simply by the fact that the indicators are also ternary. So, further enquiries are warranted in this respect.

5.6.1. Aliasing

Prior to further investigation of the results, we had to deal with the problem known as *aliasing* [Chickering and Heckerman, 1997]. Aliasing is a term coined to address the meaninglessness of states of hidden nodes parameterised with the ML estimation. The situation, that the labels of the states are at first sight meaningless, takes place because the estimates of the probabilistic parameters around the ML values are invariant no matter what the assignments of these labels are. The labelling of the states of hidden nodes is thus irrelevant to the loglikelihood value. The same occurs also with the MAP estimation when the prior parameters are uniform.

Therefore, we have relabelled the states of the latent constructs after the EM estimation was completed on the basis of indicator variables and their predictions on latent variables. For each latent construct, we have achieved this by instantiating all the indicator variables to one and the same state and calculating the marginal probability distribution for the latent variable given this state. The state of the latent construct that received the most probability mass has received the same label as the state to which the indicators were instantiated. This procedure was repeated for all the three states of the indicator variables. It is worth reporting that we had no troubles in assigning right labels to latent constructs, because for each instantiated state of indicators, at each time some different state of the latent variables received at least 99% probability mass.

5.7. Results of comparing the competing models

We have computed the relative posterior probability of the alternative models using both the CS and the BIC variants of the approximation. These measures do not suffer from a deficiency of the model's maximum likelihood function that in general grows along with the number of model parameters reaching its maximum for the saturated (fully interconnected) model. In contrast, the marginal likelihood function strikes a balance between the number of parameters and the likelihood and can be considered an objective test of the goodness of fit of a model since it is not dependent on the number of parameters.

Although we postulate a priori, i.e., before seeing any data, that Model 1 is most likely, we will let the data alone determine the posterior probability of the models. As a result, we specify the uniform priors on the five different model structures.

As regards the prior parameters on the models' probabilities, we haven't imposed any particular prior distribution on them. More precisely, we opted for the uniform Dirichlet distribution with hyperparameters $\alpha_{ijk} = 1$, for all i, j, and k. As a result of this uninformed prior distribution, we let the data alone determine the entries in the conditional probability tables. Furthermore, we assume that all four models are equally likely *a priori*. To find the maximum likelihood estimates of the parameters we have used the EM algorithm taking the highest maximum from among 100 runs with different random initial parameter values.

Four indicators for Trust and five for Involvement

In our first experiment we carried out estimation of the marginal likelihoods of the models 1-5 with four indicators for Trust, and five indicators for Involvement. The following table contains the CS and BIC approximations of the marginal likelihoods for the considered models, and the Bayes factor between CS scores of Model 1 and other models. This experiment had this shortcoming that the BIC and CS scores were calculated with the assumption that the effective dimension is equal to the structural dimension, which might not be true. This was necessitated by the fact that we were not able to compute the effective dimension of the model due to the computational complexity of this task. We note that the higher the value of logarithm, the higher the likelihood of the model. The approximations of the marginal likelihood values are given for each considered model in Table 5.7.1.

Rank	**Model**	**CS**	**BIC**	**Bayes Factor**
1.	Model 1	**-3656.4822**	**-3371.8602**	1
2.	Model 5	-3658.7878	-3372.0495	10.030
3.	Model 2	-3665.9296	-3386.7782	1.2675E+04
4.	Model 4	-3670.0782	-3386.8895	8.0291E+05
5.	Model 3	-3688.8202	-3411.2048	1.1072E+14

Table 5.7.1 The CS and BIC approximations of the marginal loglikelhoods of the model.

To interpret the results reported in Table 5.7.1, we will recall that the posterior probability of any Bayesian network model structure after seeing the data $p(B_s|D)$ is by Bayes' rule proportional to prior probability of the model $p(B_s)$ and the marginal likelihood $p(D|B_s)$:

$$p(B_s \mid D) \propto p(B_s)p(D \mid B_s). \qquad (5.11)$$

When we assume that all the alternative models are a priori equally likely, i.e. $p(B_s)$ is uniform, then in order to determine which model structure is most probable *a posteriori* we can resort to the value of the marginal likelihoods $p(D|B_s)$.

We have found out that the model with the highest posterior probability from the models we deliberate is Model 1. Looking at the Bayes factor we can see that this model is 10 times more likely than the second best model, i.e., the Model 5. Models 2 and 4 are almost equally likely, but Model 2 is about 63 times more likely than Model 4. Model 3 is the least likely model and is as much as 1.1072E+14 times less likely than the best Model 1.

Three indicators per construct

In the following step, we have performed another experiment with only three indicator variables per latent construct resulting in the models with 10 nodes instead of 13 nodes. The rationale behind this experiment was to exclude the case in which the results in the previous study were affected by the possibly incorrect value of the effective dimension, and with the smaller number of variables we expected to overcome the difficulties in the calculation of the effective dimension. The reduction of variables was achieved by taking only three indicator variables per each construct that we treat as latent, i.e. Trust and Involvement. The particular indicators that we have chosen in this experiment were those for which the measurement models scored best. We note that from among the measurement models with three indicators, Model 2 for Trust (see Table 5.5.1) and Model 7 for Involvement (Table 5.5.2) are the most likely. Consequently, as indicators of Trust we have used the variables Tr2, Tr3, and Tr4, whereas the variables Inv3, Inv4, and Inv5 were chosen as the indicators of Involvement.

We were able to compute the effective dimension for all the four alternative models with reduced number of indicator variables using again the method proposed by Settimi and Smith [1999]. We found out that the effective dimension for each model was the same as the model's structural dimension. This outcome suggests that also the quantity of effective dimension for the models with the original number of indicators is not lower than the number of structural parameters, however we do not know of any method to verify this conjecture theoretically.

For the experiment with three effect indicators per construct, we have assumed first that the priors α_{ijk} are uniform with respect to the combinations of the parents $\alpha_{ijk}=\alpha^*p(\mathrm{x}_i, \mathrm{pa}_i|B_s)$ with the equivalent sample size α of 1. The results of the experiment are presented in Table 5.7.2. Here, again Model 1 is ranked first *ex aequo* with Model 5 as the most likely model. This implies that the direction of the relationship between Involvement and Trust is not so important.

Rank	Model	Eff. Dim.	CS	BIC	Bayes Factor
1.	Model 5	80	-2663.2013	-2494.8391	1
2.	Model 1	80	-2663.2248	-2494.8522	1.024
3.	Model 4	68	-2680.2940	-2498.6335	2.650E+07
4.	Model 2	68	-2681.3855	-2498.7087	7.894E+07
5.	Model 3	64	-2702.6436	-2517.5341	1.348E+17

Table 5.7.2 The aproximations of the marginal likelihoods for models with three indicators.

Following, we let the priors be again uniform but with hyperpriors of 1 for all i, j, and k. These hyperpriors were chosen in order to test sensitivity of the marginal likelihood to different priors, and not as a form of prior knowledge. The results collected in Table 5.7.3 indicate now that the most probable model structure is encoded with Model 2, followed by Model 4. We can infer thus that the parameter priors have an impact on the posterior probability of the model. This is probably because there is not much data, and especially there are no data for the hidden nodes.

Rank	Model	Eff. Dim.	CS	BIC	Bayes Factor
1.	Model 2	68	-2758.4438	-2504.9746	1
2.	Model 4	68	-2761.8333	-2504.9477	29.6511
3.	Model 1	80	-2762.8888	-2505.9039	85.1999
4.	Model 5	80	-2771.9808	-2505.6981	7.569E+05
5.	Model 3	64	-2772.0263	-2523.2153	7.921E+05

Table 5.7.3 Ranking of the hypothetical models in the as assuming priors on parameters α_{ijk}=1.

Combining domain knowledge with data

In the above experiments, we have let the priors be uninformed. In this way, we showed that we do not need to use potential prior knowledge on the nature of relationships into the estimation of probabilities, and the posterior probability of models. In other studies, a situation can take place that the researcher has some prior domain knowledge, or some elements of the established theory, that they find important to introduce into the model in focus; the priors in this situation would reflect this knowledge. Besides assigning prior parameters to specific probabilities in the CPT's, it is easy in the Bayesian network framework to assign prior probabilities both to specific parent-child links and to whole network structures. Due to space limitations, we do not consider such scenarios here. This combination of prior knowledge with data is the essence of Bayesian modelling and is closer to what some authors argue, that model building should rely more on accumulated knowledge for improved decision support [Ehrenberg, 1994]. Heckerman *et al.* [1995] derive formula for a Bayesian scoring metric that enables incorporation of prior information in the form of probabilities of some links.

5.8. Marginal probabilities, structural and measurement relationships

Besides the statements of presence or absence of direct relationships between the latent constructs implied by the most likely model, i.e., Model 1, we should examine the quantitative aspect of the models by looking at the conditional probability tables, and by considering the strengths of dependencies among the variables in the most likely model. Of course, we have to keep in mind that the labels of states for Trust and Involvement have been determined *a posteriori* on the basis of influence of indicators as we reported in section on aliasing.

Furthermore, we could perform other analysis with the most likely model, including what-if simulations, forward, backward and inter-causal inference as well as classification and prediction. These operations can be accomplished in a similar way like we have demonstrated in the previous study, so in order to keep the contents clear and concise, we refrain from doing this analysis in the current study.

5.8.1. Marginal probabilities

Let' start with the marginal distributions of the four variables shown in Table 5.8.1.

	Satisfaction	Trust	Involvement	Loyalty
low	0.054	0.143	0.279	0.083
moderate	0.413	0.426	0.478	0.386
high	0.532	0.430	0.242	0.530

Table 5.8.1 Prior marginal probabilities.

We can see that there are small marginal probabilities of low scores for Satisfaction and Loyalty. The most probability is that the respondents experience high Satisfaction, and also high level of Loyalty.

For Trust, the table shows equal probabilities for the moderate and high level. We can also see higher probability of low Trust than we observe for Satisfaction and Loyalty. We must remember that Trust is modelled as a latent construct, whereas Satisfaction is treated as an observed variable.

Marginals for Involvement are even smoother than for Trust and other variables, i.e., they are closer to the uniform distribution; this could be the effect of the relatively flat conditional probabilities of Involvement given Trust, which could be a result of the EM estimation. Of course, we cannot determinedly conclude whether these marginals are far from the true scores.

5.8.2. Structural relationships

The structural relationships, i.e., the relationships between the latent constructs are in our approach described by means of the conditional probability tables. It would be interesting to review the CPT's of the most likely model found with all five and four indicators for Involvement and Trust, respectively. We can consider the CPT's also as a measure of validation of the latent construct model. If the tables appear to be theoretically reasonable, then it should be seen as evidence in favour of our methodology.

Table 5.8.2 shows conditional distributions of Trust conditional on Satisfaction. In the row labelled "Counts", we show here in fact the *expected* number of respondents N'_{jk} for a given configuration of parents' states estimated by the EM algorithm (see Expression 5.6). As was the case with observed counts in Section 4.4.5, this number should be taken into consideration to assess significance of the conditional probabilities in the configuration. We can notice a strong positive effect of Satisfaction: the higher probability of Satisfaction, the smaller probability of low Trust and the higher the chance of high Trust. Accordingly, we can conclude that the nature of this relationship is theoretically appealing. We should however be cautious with the exact interpretation of these probabilities without any further investigation of the effect of the EM algorithm; to be more precise, we should rather view them as an indication of the strength of the relationship or of the most likely quantitative nature of the dependency.

Satisfaction	low	mod	high
Counts	21.7	171.9	222.3
low	0.680	0.170	0.068
moderate	0.258	0.617	0.294
high	0.062	0.213	0.637

Table 5.8.2 Conditional probabilities for Trust given Satisfaction.

Next, let us inspect the most interesting observations found in the conditional probability table of Involvement given Satisfaction and Trust shown in Table 5.8.3. We can see, for instance, that given low Trust, probability of low Involvement, varying from 0.727 to 0.876, is almost independent of Satisfaction. In other words, given low Trust, it does not make any difference whether a customer is satisfied or not in order to be involved. On the other hand, given low Satisfaction, the probability of low Involvement decreases from 0.815 to 0.446 as Trust grows (the probability 0.371 given high Trust is not reliable because of the low expected count).

Satisfaction	low			mod			high		
Trust	low	mod	high	low	mod	high	low	mod	high
Counts	5.8	15.8	0.001	109.5	29.9	32.4	72.2	16.1	134.0
low	0.815	0.446	0.371	0.876	0.339	0.098	0.727	0.188	0.088
moderate	0.107	0.434	0.345	0.093	0.638	0.542	0.185	0.716	0.385
high	0.078	0.120	0.283	0.032	0.023	0.359	0.089	0.096	0.526

Table 5.8.3 Conditional probabilites for Involvement.

Finally, conditional distributions for Loyalty given Involvement and Trust are presented in Table 5.8.4. Let us focus on the effect of high levels of customer Involvement and Trust apart. First, we can see that given high Involvement, the probabilities of high Loyalty vary from 0.285 to 0.902 depending on the level of Trust; second, given high Trust, the probabilities of high Loyalty vary, depending on Involvement, from 0.579 to 0.902. This suggests that high Trust has stronger effect on high Loyalty than high Involvement has.

Involvement	low			mod			high		
Trust	low	mod	high	low	mod	high	low	mod	high
Counts	3.59	0	80.0	51.4	56.4	11.0	132.5	5.48	75.4
low	0.405	0.068	0.137	0.146	0.023	0.022	0.264	0.103	0.023
moderate	0.402	0.730	0.284	0.523	0.535	0.281	0.451	0.375	0.075
High	0.194	0.202	0.579	0.331	0.442	0.697	0.285	0.521	0.902

Table 5.8.4 Conditional probabilites for Loyalty.

In conclusion on the nature of the structural relationships, we can infer that the conditional probabilities relating to the latent constructs seem to be theoretically and practically appealing. Consequently, we can also conclude that our proposal of handling latent construct by local Naïve Bayes models seems to perform well. However, we need to stress that further studies involving the four constructs in focus are needed and more support in the literature should be found in order to corroborate this statement.

5.8.3. Measurement relationships

In this section, we report some of the conditional probability tables that contain parameters of the relationships between latent constructs and their respective observed variables.

For instance, let us interpret the probabilities in dependency of the indicator variable Tr1 given the latent construct Trust shown in Table 5.8.5. We see that these conditional probabilities are reasonable, i.e., the probability of low response on the indicator Tr1 ("the company shows appreciation for me as customer") given low trust in general is 0.669, the probability of moderate response given moderate trust in general is 0.721, and the probability of high response on the indicator Tr1 given high trust in general is 0.762.

Trust	low	mod	high
Counts	187.6	61.9	166.4
low	0.669	0.038	0.021
moderate	0.307	0.721	0.216
high	0.024	0.241	0.762

Table 5.8.5 Conditional probabilities for the indicator Tr1 given the latent construct Trust.

Involvement	low	mod	high
Counts	83.6	118.9	213.4
low	0.808	0.037	0.031
moderate	0.137	0.893	0.061
high	0.054	0.069	0.906

Table 5.8.6 Conditional probabilities for the indicator Inv3 given the latent construct Involvement.

As another example, we will take the CPT of the observed variable Inv3 ("I feel proud to be a customer of ...") as an indicator of Involvement (Table 5.8.6). We can see that the probabilities that define the measurement power of this indicator are higher for the corresponding states of the two variables in comparison to the indicator Tr1, as the probability of low response on the indicator Inv3 given low true involvement amounts to 0.808, the probability of moderate response given moderate true involvement in general is 0.893, and the probability of high response on the indicator Inv3 given high involvement in general is 0.906. We can interpret that Inv3 is a more reliable indicator than Tr1 because it explains the respective latent construct in a way that reduces the uncertainty. We could also say that the error of measurement in case of the indicator Tr1 is greater than in case of Inv3. This latter finding can be interpreted as an disadvantage of our approach, since it does not specify the error of measurement as an explicit component of the measurement model, so it is not possible to distinguish what is the measuring power of the indicator (in terms of "loading" on the latent construct) and what is the measurement error. Other theories than the classical true-score theory of measurement [Lord and Novick, 1968] could provide some insight and theoretical background for our approach. The remaining tables of indicator variables are included for the consultation in Appendix D.

5.9. Latent constructs as averaged items

As we have mentioned earlier, to date, there exists no standard approach to dealing with latent constructs and to accounting for measurement models in the Bayesian network literature. In all encountered studies we have found that the authors use one-item measurement scales and treat the observed variable as if it was the latent construct itself [e.g., Anderson and Lenz, 2001]. This approach does not allow for accounting for the measurement error, does not allow for assessment of the validity and reliability of the scale, and, as such, is not recommended [Bagozzi, 1994a; Dillon *et al.*, 1997]. Another potential approach is to use multiple-item scales as measurements and create an index for each of the latent constructs on the basis of the indicators prior to further analysis; typically one takes the average over the indicators. The resulting variable is used as an observed variable in the model. This approach is employed in linear regression modelling [Gefen *et al.*, 2000].

In this section, we construct a model for our domain taking this last approach, and compare it with the model with latent construct as we proposed in the previous section. The objective of this study is to find out whether taking the average from the indicator variables to create a new variable gives the same results with respect to 1) relationships between variables, and 2) classification accuracy.

5.9.1. Study design

To start with, we have created new variables representing the latent variables by taking the average of the respective observed indicator variables. The value of Involvement for each respondent was obtained by taking the arithmetic mean over original values of variables Inv1, Inv2, Inv3, Inv4 and Inv5. Similarly, variables Tr1, Tr2, Tr3 and Tr4 were used in the same way to create the variable Trust. After these two new variables were constructed, we have them recoded by discretising their values into three new states using the following aggregation: values from 1.0 to 4.99 received the meaning "low", values between 5.0 and 7.99 were labelled "moderate", and values from 8.0 to 10 received the meaning "high". Loyalty and Satisfaction were also assumed to be ternary.

The structure of the model is the same as presented in Figure 5.3.1a; it is the most likely model structure among these we investigated in the previous section. We will refer to the model with latent variables replaced with a variable that was the average of the indicator variables as Model II, and to the model with explicit handling of latent construct as Model I. Because there were missing data in a data set, the probability tables for the model were obtained by taking maximum likelihood estimates using the EM algorithm from 100 different initialisations.

5.9.2. Results

In the first step, we compare both approaches by means of classification accuracy. The classification task consists in predicting the state of loyalty (Loy) based on the values of the remaining variables. The actual observed state of Loyalty for each observation has been compared with the state obtained from the inference in the model. The candidate models have been benchmarked by means of various scores expressing precision of the classification, including the prediction accuracy, the log loss and the Brier score [Gaag and Renooij, 2001; Panofsky and Brier, 1968].

	Model I	**Model II**
Cases verified	389	390
Cases classified correctly	263	249
Classification accuracy	67.6 %	63.8 %
Log loss	0.6920	0.7230
Brier score	0.1831	0.1948

Table 5.9.1 Quality of classification.

From the results in Table 5.9.1 it follows that Model I is a slightly better classifier than Model II since it has higher accuracy and lower log loss and Brier scores.

In Table 5.9.2, we present prior marginal probabilities of the four constructs. As we could expect, probabilities of Satisfaction and Loyalty are very alike compared to those in Table 5.8.1. As regards Involvement and Trust, we can see that the EM algorithm applied in model I is responsible for softening of the "rough" probabilities in case of Model II.

	Satisfaction	Trust	Involvement	Loyalty
low	0.052	0.088	0.346	0.007
moderate	0.414	0.571	0.507	0.387
high	0.534	0.340	0.146	0.537

Table 5.9.2 Prior marginal probabilities.

Next, the two models can be compared also by means of the conditional probability tables. In Table 5.9.3 we show the conditional probability table for Trust given Satisfaction.

Satisfaction	low	mod	high
Counts	21.7	172.6	222.6
low	0.581	0.094	0.035
moderate	0.327	0.692	0.502
high	0.092	0.214	0.463

Table 5.9.3 Conditional probabilities of Trust given Satisfaction for Model II.

Firstly, Table 5.9.3 shows that the conditional probabilities are close to those in Table 5.8.2. This observation is in favour of the labelling as effect of aliasing. The only big difference is in case of moderate and high Trust given high Satisfaction.

Satisfaction	low			mod			high		
Trust	low	mod	high	low	mod	high	low	mod	high
Counts	12.6	7.1	2	16.22	119.4	37	7.8	111.8	103
low	0.997	0.577	1	0.907	0.456	0.194	0.808	0.284	0.106
moderate	0.003	0.423	0	0.092	0.543	0.666	0.191	0.652	0.417

high	0	0	0	0.001	0.001	0.138	0.001	0.062	0.475

Table 5.9.4 Conditional probabilites for Involvement for Model II.

Comparing conditional probabilities for Involvement in Table 5.9.4 and 5.8.3, we can see first of all differences in the expected counts. Furthermore, the EM algorithm applied on the latent construct model results in probabilities that are softer and smoother than in the model with averaged items; in general, the shape of each distribution is conserved.

Lastly, we consider the CPT for Trust shown in Table 5.9.5. Again, we can see big differences in the expected counts between the two methods. For instance, there appears to be no observations in the configuration in which Involvement is high and Trust is low while in the latent construct model (Table 5.8.4) there are expected to be 132.5 cases in this combination. As regards the probabilities, the only remarkable discrepancy takes place for moderate Involvement and Trust.

Involvement	low			mod			high		
Trust	low	mod	high	low	mod	high	low	mod	high
Counts	33.65	90.4	20.19	3	140.9	67.6	0	7	54.13
low	0.511	0.098	0.001	0.0	0.030	0.014	0.333	0.001	0.001
moderate	0.383	0.612	0.499	0.666	0.394	0.313	0.333	0.166	0.058
high	0.104	0.289	0.499	0.333	0.574	0.671	0.333	0.833	0.941

Table 5.9.5 Conditional probabilites for Loyalty for Model II.

In conclusion, firstly, the latent construct model outperforms the "standard" approach in classification accuracy. Taking the average is not the optimal technique because it ignores the relative importance of the indicator variables in measuring the abstract concept. So, there is potential loss of valuable information. This loss can also be seen by interpreting the classification accuracy, where averaging of indicators for Trust and Involvement results in worse classification function for Loyalty.

Secondly, there exist only slight differences in conditional distributions. Conditional distributions are sharper, with in case of the "averaged items" approach, whereas for the latent construct model they tend to be softer and more alike the uniform distributions.

Thirdly, unlike the "averaged items" approach, the latent construct model allows for assessing the validity of the scale, as we have shown in Section 5.5.1.

5.10. Conclusions and future research

5.10.1. Conclusions

Our aim in this chapter was to investigate further the research question number 1, namely, how marketing theories can be discovered by means of the Bayesian network approach.

In particular, in relation to each goal and sub-goal defined in Section 5.1 in this chapter, the case study delivered the following conclusions:

1.a. We evaluated Bayesian networks in the deductive CS&L research.

In our deductive approach, we postulated five hypothetical models of Customer Satisfaction and Loyalty. On the basis of speculations about a theory of CS&L, we are able to construct a theoretical model, and validate it against the empirical data.

Taking the Bayesian score as a measure of the goodness of fit, we can validate a hypothesis of presence or absence of a direct relationship between two constructs.

From the results, we can see that both the Cheeseman-Stutz and the Bayesian Information Criterion scores are highest for the first model. We can therefore conclude that the most probable model among the 5 analysed models is Model 1. This model suggests existence of a direct dependence of Involvement on Satisfaction and Trust. This dependence is more probable than the dependence of Involvement on Satisfaction only. Model 1 has been postulated in fact by the marketing research company before seeing any data.

In conclusion on the deductive research with Bayesian networks, we argue that the deductive approach can be successfully carried out within the Bayesian network modelling; it must be remembered however that, unlike it is the case with other techniques applied in CS&L research, it is not possible with the presented approach to perform validation based on the marginal likelihood to strictly confirm whether the model can be accepted or should be rejected.

1.b.i. We have proposed and evaluated new methods for handling of latent constructs and accounting for the measurement model in BN modelling.

Nowadays, no method exists, to our knowledge, of incorporating the structural and measurement models explicitly into the Bayesian network modelling. Therefore, our method of accounting for these two models in one holistic analysis can be seen as a contribution of a great importance. First of all, the results of our proposed method are theoretically sound in the sense that structural models that we a priori assume more likely, indeed score higher. It is apparent that the proposed method of handling the structural model can be used to test the presence or absence of some theoretical relationships between latent constructs. Furthermore, conditional probabilities between latent constructs, i.e., defining the structural model, are meaningful and provide valuable insight into the nature of relationships. Additionally, the relationships in the measurement model are also meaningful and show that the approach, which we proposed in this chapter, is valuable and performs well.

Furthermore, we have performed comparison with the "standard" approach in which latent constructs are not treated as latent but are constructed as the average over indicator variables. We have found that our proposed latent construct model outperforms this standard approach in the classification accuracy. Taking the average is not the optimal technique probably because it ignores the relative importance of the indicator variables in measuring the abstract concept. So, there is potential loss of valuable information. This loss can also be seen by interpreting the classification accuracy, where averaging of indicators for Trust and Involvement results in worse classification function for Loyalty. Secondly, there exist only slight differences in conditional distributions. Conditional distributions are sharper in case of the "averaged items" approach, whereas for the latent construct model they tend to be softer and more alike the uniform distributions. Thirdly, unlike the "averaged items" approach, the latent construct model allows for assessing the validity of the scale.

In summary on the latent construct modelling approach by means of local Naïve Bayes, we conclude that our approach performs very well. Our method proves to be useful and shows the added value of this work, but further investigation should be carried out.

1.b.ii. We have proposed and evaluated a method for validation of latent constructs within Bayesian network modelling.

To be precise, the proposed method enables validation of the measurement instrument to the extent that the effect indicators are related either to one latent construct or to two potential latent constructs. It enables testing which items, in sets of two, three or four items, relate collectively to one latent construct. For two constructs on which we have applied our method, we have found that all four and five indicators are most likely common indicators of one construct, respectively. We have found that our method could also be used for discovery or validation of multidimensional nature of latent constructs.

Furthermore, we have compared the results of the proposed method with the classical methodology applied for this purpose, i.e., factor analysis and reliability analysis (Cronbach's alpha). We found that the scales used are indeed unidimensional and have high reliability coefficients. Thereby, the results of both approaches are fully consistent with each other, which suggests that the method of construct validation that we propose is, at least for the data at hand, externally valid and performs well. Naturally, further applications and validation of the technique with diverse data should be undertaken to corroborate its value.

1.b.iii. We have proposed and evaluated a method for finding the dimensionality of latent constructs in Bayesian network models.

We have defined dimensionality as a number of states that a latent construct most probably takes. It can be important for the theory under scrutiny because it might happen that depending on the dimensionality of the construct, the marginal likelihood of the entire model can be different. It can also be useful with respect to the construct itself, because it shows the scale on which the construct operates. For instance, if we conceptualise satisfaction, we could find out whether it is a dichotomous variable, and takes only two states "low satisfaction", and "high satisfaction", or it spans rather over more intermediate values, e.g., "low satisfaction", "moderate satisfaction" and "high satisfaction". For two constructs for which we have applied this technique, we have found that both concepts are best represented as ternary variables.

2. With regard to the added value of modelling marketing problems with Bayesian networks, we show and illustrate the potential of combination of prior knowledge with data at hand.

By prior knowledge we mean our beliefs, or theoretical insights, concerning character of specific conditional distributions for each combinations of a focal construct's parents' values. These prior beliefs are then faced with observational data from our study to determine the posterior estimates of the probabilities defining these conditional distributions.

We have designed two experiments in which we imposed different priors on parameters of these local distributions. These priors can be seen as "uninformed" in the sense that they do not represent any concrete prior knowledge: the two models examined in these experiments were different from each other in the amount of our ignorance. We observed that they have indeed en effect on the posterior distributions, and even on the marginal likelihood of the model. In our experiment we have found that these priors, even more importantly, have an effect on the relative probability between models. This is probably because there is not much data, and especially there are no data for the hidden nodes.

In conclusion, we must note that this kind of introducing prior knowledge into the development of theory of phenomenon under study is characteristic of the Bayesian data analysis. This type of analysis can be especially useful when important accumulated knowledge exists with respect to the specific character of the relationship, that we want to account for, between two adjacent constructs, or when data at hand are scarce, or when data come from sources of different kinds.

3. Whenever appropriate, we have pinpointed the strengths and weaknesses of Bayesian network in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc.

In the course of the discussion in this chapter, we have identified the potential of determining the values of latent constructs, and testing for omitted constructs as the strengths of the Bayesian network approach.

Most importantly, we found in this case study that the approach of handling latent constructs that we proposed provides an easy possibility to determine the value of the latent construct on account of its indicators. This can be achieved for all cases in a data set, even for these cases for which some indicator variables have missing values.

A direct consequence of the work in this chapter is that we can check out whether introducing new latent constructs does not increase the likelihood of the model. As an example, let us assume that we have constructed a theoretical model of CS&L loyalty for which we obtained a specific value of the posterior (approximated) probability. Now, we could introduce another construct into this model, by positioning it in the model in a place implied by the conceptual meaning of this construct and our new theoretical hypothesis concerning it; the Bayesian network approach enables calculating the posterior (approximated) probability of this new model. Of course, higher probability implies now that our new theory is more likely than the old one, whereas, accordingly, smaller probability will imply that the new theory is less likely. We note that we have not examined such scenarios in this work, but we recognise such a potential of Bayesian networks with latent constructs.

Furthermore, the Bayesian network approach with latent constructs is subject to weaknesses, including problems with estimation, calculation of effective dimension, its inability to control for the measurement error, and inability to undergo strict confirmation. Handling of latent constructs and measurement model is still under examination, and is the focus of active research at the moment. One weakness that we must realize when applying Bayesian networks with measurement models is that we must use approximations of the marginal likelihoods. These approximations require that we estimate conditional probabilities with the EM algorithm, so all consequences of the use of this algorithm must be also taken into account. An important issue that must be mentioned here is the potential problem of under-identification. More precisely, there is no guarantee, with the Bayesian network approach with latent constructs, of finding the global optimum for model parameters (conditional probabilities); we have not done any investigations in this direction, so we stay cautious with making firm statements about this issue. We are confident that the problem of over-identification is, unlike with the SEM approach, nonexistent, which is a advantage in favour of the BN approach. Furthermore, the requirement of multiple restart of the EM algorithm, or slow convergence, can be seen by some authors as another weakness, although in our opinion this disadvantage can be quite well resolved by methods proposed in the Bayesian network literature.

Also the calculation of the effective dimension for latent construct models should be recognised as a weakness, since this calculation cannot be performed in every model. In particular, the more variables are treated as true latent constructs in the model, the more difficult it is to obtain the effective dimension.

One of the major weaknesses of the Bayesian network methodology nowadays is that there does not exist any established method for structural and measurement modelling. The methods that we propose in this chapter are attempts to solve this problem. However, a major drawback of our method of measurement modelling is that it is still not able to control for measurement error.

We have also observed in this case study that the posterior probability of the models as the goodness-of-fit measure can be viewed as a weakness in the sense that it does not enable categorical confirmation of the model. Typically in deductive research, the aim of building a theoretical model is to test it empirically to find evidence as to accept or reject this hypothesized model. This can be termed the *strict confirmatory* modelling [Joreskog and Sorbom, 1993]. Such a procedure is not feasible by taking the Bayesian network approach. To be precise, it is not possible to confirm a theoretical model in the strict sense, as the marginal likelihood measure, until today, cannot be treated with some form of statistical significance test. Furthermore, to the best of our knowledge, no statistical tests have been proposed for significance in the difference between the marginal likelihoods of two various models. Nevertheless, it must be noted that a Bayesian network model can be empirically validated in the strict sense using the constraint-based approach [e.g., Spirtes *et al.*, 2001].

5.10.2. Implications

From research presented in this case study we can draw implications both for researchers engaged in basic research on Customer Satisfaction and Loyalty, as well as for practitioners involved in applied e-loyalty modeling.

5.10.2.1. Implications for research

Again, as we found in Chapter 4, we postulate here that the Bayesian network approach makes theoretically sound inference from data. In particular, the results of the customer loyalty study in this chapter corroborate the *a priori* postulated theoretical model of this phenomenon. This suggests that the model validation procedure based on the posterior probability of the model is a valuable way both of discovering and corroborating the theory of the Customer Satisfaction and Loyalty.

We have proposed and examined a method of incorporating the measurement model into causal modelling with Bayesian networks by introducing latent variables operationalized with multi-item measurement scales directly in the model. In particular, we encourage CS&L researchers to apply the proposed approach in their research practice, as our experience delivers very positive results on our approach. Furthermore, we suggest to get familiarised with the method since it enables performing construct validation and finding the best dimensionality of latent constructs. The procedure of construct validation taken in this study aims to assess whether the indicator variables relate to one potential construct, or to more constructs. In our method, we consider dimensionality as the most likely number of values that a latent construct takes on. Moreover, we have found that aliasing does not pose any problem, since the meaning of states of latent constructs can easily be established from the indicators. All in all, the results in all these issues are very constructive but require further examination.

Again, we have demonstrated that missing data pose no problem for the proposed methodology when estimating the parameters of the model. By means of the EM algorithm, missing values in the network can be imputed in a very sound way by using all the knowledge, or theory, that the model represents. As a way of example, even when a particular respondent has responded to one question in a survey, it is very easy to make use of this single datum, and to estimate the most likely values of other variables for this respondent (by means of reasoning in the model); naturally, added value of this particular case in the model estimation is typically negligible, but by this example we would like to point that missing data poses no problem. This is an interesting implication for researchers faced with bad quality data since often they are forced to leave out the cases with missing values, which can contribute to less powerful tests of significance and impair the quality of their work. Furthermore, even when the model is ready to use, it is perfectly feasible to adapt this existing model in the light of new data.

Simultaneously with the procedure of construct validation, we can check and discover whether introducing new latent variables does not suggest existence of new, or omitted, latent constructs. In that case, if the network structure augmented by the introduction of a latent construct (of course without their equivalent indicator variables) would represent higher value of the likelihood, then this might be an indication that this new, previously not considered construct, can potentially play an important role in the theory under consideration. By looking at relationships between this new construct and the remaining constructs, we can also get an idea what omitted concept the construct should represent [see e.g., Heckerman *et al.*, 1999].

Our implementation of the presented deductive approach proved suitable with the use of Bayesian networks with latent constructs. Given its potential to locate "unknown" latent constructs, we propose that the Bayesian network approach with latent constructs is especially suitable for explanatory and exploratory research, and in the further instance for confirmatory. Furthermore, we are quite convinced that even when all concepts are measured with one-item measures, the approach can be found useful.

5.10.2.2. Managerial implications

Although this chapter is fully devoted to validation in theoretical research, some recommendations for marketing managers concerning customer involvement and loyalty can also be drawn. For instance, from our finding that given high Trust there is more probability of high Loyalty than given high Involvement (this effect is stronger), we can recommend that the companies should stimulate high confidence of their customers rather than their engagement.

Next, we argue that practitioners will find the presented approach valuable, as unlike it is the case with other techniques, it easily enables to determine the value of the latent construct based on the values of the indicator variables. As a result, they can perform simulations by assuming some values of the observed variables, introducing this information as evidence into the model, and by performing reasoning in the network they can find out the posterior distribution for the corresponding latent variables; even more interestingly, they can see the effect of these assumed values of indicators on other constructs in the network. We believe that this capability is of great value to marketing managers.

Another important implication of this research for marketing managers can be that they will find the use of latent construct Bayesian network models easy and intuitive. They should find it easy to advance several competing structural models, link the latent constructs to their indicators, and draw conclusions from comparison between these models. This finding should yet be corroborated in practice by exposing our approach to managers and marketing practitioners.

5.10.3. Limitations

We must note a few limitations of research presented in this case study.

First of all, we should take into account that we have not performed any thorough investigation of the quality of the data in relation to the reliability and validity of the scales used and the measurements. This concerns for instance issues such as the convergent and discriminant construct validity. Specifically, we should note that possibly many different techniques should be typically applied to establish a satisfying level of confidence in the reliability and validity of the data.

Measurement modelling has been originally developed as an instrument of accounting for the measurement error, which should be the explicit component of marketing models [Steenkamp and Baumgartner, 2000]. In the classical true-score theory of measurement [Lord and Novick, 1968], the observed score equals the true unobserved value plus the error term. From the point of view of this theory, the measurement modelling approach that we presented in this case study can be criticised for departure from this principle of full incorporation of the measurement error in the holistic analysis. A limitation of the proposed approach to measurement modelling could thus be that the measurement error in the relationships between latent construct and the corresponding observed variable cannot be separated qualitatively from the true score for the latent variable. In our approach, this error manifests itself rather in the conditional distribution for the observed variable given the true score on the latent construct, and more precisely in the uncertainty around the corresponding state of the indicator; we can say that the more uncertainty, i.e., the more probability mass is distributed to other states of the indicator, the greater the measurement error.

We can conclude that the handling of the structural model as consisting of hidden nodes estimated by the EM algorithm leads to theoretically sound conditional probabilities. However, we do not know exactly what the precise impact is of the EM estimation on the conditional probabilities. We conjecture that conditionals are likely to be too soft and the EM estimation makes the distribution be smoothed compared to "true" conditional probabilities between latent constructs in reality, so they should be taken with caution.

While discussing construct validation method, we considered existence of only two latent constructs that the indicators could relate to. Therefore, settings in which three and more constructs are present should be tested.

A potential serious bottleneck of modelling hidden variable Bayesian networks is the calculation of the effective dimension, which is required to approximate the marginal likelihood of the model. Taking the structural dimension on the other hand can bias the results. The problem of calculation of the effective dimension grows with the number of hidden variables. The more hidden variables there are in the model, the more time it takes to estimate it. Furthermore, another limitation of the approach with latent constructs is that no precise measures of marginal likelihood of the model exist, so that one has to fall back on approximations, such as BIC and CS that can be not precise.

Last but not least, the theoretical model in this chapter should be again seen more as an illustration of our Bayesian network approach to methodological topics addressed here.

5.10.4. Future research

We stress that further application of the Bayesian network approach with latent constructs in other customer satisfaction and loyalty settings involving diverse data sets is recommended. The recognition of Bayesian networks as a fully legitimate techniques for theoretical modelling requires that issues like reliability and validity are fully taken account of and attainable within the scope of the technique. Of these topics, in this chapter we have presented a possible measure for assessing construct validity, but this and other topics in these respects call for more attention.

One of the most important suggestions for future examination is analysis of the behaviour of the EM estimation on the conditional probabilities and marginals. It would be very interesting to carry out studies on simulated data.

Further extensions of the presented work are also possible and are absolutely worth investigating. One of the most significant topics for further exploration is the analysis of statistical characteristic and behaviour of the presented method of construct validation. For example, in our procedure of construct validation in Section 5.5.1, we performed validation of each measurement instrument in isolation from the complete model. However, the validation of the instruments could also be achieved by considering them in the broader contexts of the entire model, as it could turn out that the mutual relationships in the model play a role in assessing the impact of latent constructs on the indicators.

Since we have tested the proposed approach only on two constructs, it is too few to give any solid assessment. Hence, this method should be merely seen as an initial attempt directed at developing a construct validation procedure within the Bayesian network framework. Therefore, further thorough investigation of properties of our method is necessary in follow-up studies. Various measurement instruments already validated by other authors and well established in the literature should be used as test instances. Further evaluation of this method could be based on comparison with the standard methods applied in SEM modelling, such as multitrait-multimethod (MTMM) of Campbell and Fiske [1959].

Further work is required to corroborate the correctness of the presented approach of finding dimensionality of latent constructs. Central issue is whether models that postulate three states of latent constructs could be preferred over models having other number of states than three simply by the fact that the indicators are also ternary. So, further enquiries are warranted in this respect, for instance by observing the effect of variation of the cardinality of the observed variables from two to the original value of ten.

Thanks to recent advances in structural learning of Bayesian network models from data, methods have been proposed that facilitate finding most likely models with latent constructs directly from the data by means of efficient search algorithms [e.g., Russel *et al.*, 1995; Friedman, 1998]. The common motivation for these methods is that bringing in a new variable can simplify and compact the structure of the model. As the central feature of these methods, during the search for the most likely model, it is evaluated whether there could be any potential hidden variables in the domain, i.e., variables that are not present in the observed data. Roughly speaking, this is done by hypothesising the presence of a latent variable at a certain place in the model, and if the marginal likelihood of such an augmented structure is higher than the one of the original structure, then this variable is retained in the model. Its theoretical meaning can be then guessed on the basis of the location and relationships with other constructs. Further enhancements of these approaches and corroboration of their use in the CS&L research is one of very exciting avenues for further scientific work. Other example topic could be how the presence of hidden constructs can be detected without the need of scoring the entire model.

6. Case study 3: Practical satisfaction studies

6.1. Introduction

In Chapters 4 and 5 we applied and evaluated Bayesian networks in the task of discovering and explaining the theory of Customer Satisfaction and Loyalty. As such, our findings were relevant primarily for marketing academics that aim to gain scientific understanding of the CS&L phenomenon. After Ehrenberg *et al.* [2000], we have argued that the findings could also be indirectly important for marketing practitioners. However, it occurs that the main interest of marketing practitioners lies not so much in theoretically sound conceptual models of CS&L, but in models that more directly let them support their marketing decisions. In this chapter, we focus thus specifically on problems facing marketing managers in relation with the performance of their products/services. More precisely, we investigate the role of product/service dimensions in creation of customer satisfaction.

The output of Bayesian network model is usually presented with tables containing a series of prior and posterior (conditional) probabilities. In contrast, in this study we apply the procedures of sensitivity analysis to diagnose the dependencies in a way that they are represented with algebraic functions - often resembling linear regressions - which are more familiar than numbers, i.e., conditional probabilities alone. Such a representation allows for easier interpretation of the numerical facet of dependencies, for example, by showing their strength, and providing a simple yet rich source for enquiry. The functional form of dependencies lends itself to be portrayed using informative charts and plots. The results of the analysis can be revealed with respect to prior probabilities as well as probabilities conditional on some specific assumptions of interest.

6.1.1. Objectives

Our main objective in this chapter is to adapt and examine the Bayesian network methodology in the context of practical satisfaction research. This will help us achieve the second main goal in this thesis, which is evaluation of Bayesian networks and demonstration of its added value in practical satisfaction research.

In particular we address the following research questions:

1. How can Bayesian networks be applied in a service feature/dimension importance/performance study? More specifically, we adapt and examine Bayesian networks in service dimensions analysis for:
   a. identifying the derived importance of service dimensions for overall (dis)satisfaction judgments,

b. supporting marketing decisions by means of importance/performance analysis,
c. discovering interaction effects (synergy and negation) among service dimensions.

2. What are the strengths and weaknesses of Bayesian networks in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc.

Firstly, we adapt and examine Bayesian networks for the purpose of identifying the derived importance of potential factors for overall (dis)satisfaction judgments. Our objective will be to find out which service/products dimensions are potential sources of (dis)satisfaction. To this end, we apply a procedure based on sensitivity analysis in Bayesian networks.

Secondly, our Bayesian network approach is evaluated for the potential of supporting marketing decisions by means of importance-performance analysis. The objective of this analysis is to indicate these service dimensions on which the company should focus their resources in the first place, and which dimensions are objects of possible overkill. Some of the categories that we define are: low priority, action needed, opportunities, strengths, take care, and possible overkill.

The third topic that we discuss is if and to what extent Bayesian networks can be applied for discovering of interaction effects among service dimensions. We will adapt and examine the approach in this regard.

Last but not least, we will also investigate the strengths and weaknesses of Bayesian networks in terms of specific statistical and modelling issues, for instance by allowing for optimal use of all available data in one model.

The chapter is organized as follows. First, in Section 6.2, we elaborate on sensitivity analysis in Bayesian networks and give an illustration on a dummy model of how sensitivity analysis can be carried out within the Bayesian network framework. Our approach to classification of service dimensions is the topic of Section 6.3. Explanation of importance-performance analysis is addressed in Section 6.4. Data issues are the topic in Section 6.5. In Section 6.5 we present and discuss the results. We close with conclusions in Section 6.7.

6.2. Sensitivity analysis in Bayesian networks

One of the fundamental functions of Bayesian networks is to take advantage of the efficient representation scheme of the joint probability space over the modelled system and exploit it to calculate some probabilities of interest. For example, the primary use is to retrieve a probability distribution for a node of interest, called a *target* node, conditional on some set of *query* nodes, called also *explaining* or *evidence* nodes, when their values become available. Other potential use is to find the probability of some specific configuration of a node's values.

The results of such calculations can be achieved automatically by means of probabilistic inference algorithms that are typically implemented in the Bayesian network-enabled software. The user can simply enter queries to the Bayesian network by identifying target nodes, and assigning values (states) to explaining nodes. The question that often arises in this respect is how sensitive those resultant posterior probabilities are to changes in the numerical strengths of the dependencies. This is where the issue of sensitivity analysis comes into play.

Generally speaking, sensitivity analysis in a mathematical model pertains to investigation of the effects of the inaccuracies in the model's parameters on its output. This can be achieved for instance by "brute force" variation of the model's parameters and evaluation of the change in model's output. For a Bayesian network model in particular, sensitivity analysis can be approached twofold: empirically and theoretically [Kipersztok and Wang, 2001].

6.2.1. Empirical sensitivity analysis

The empirical approach to sensitivity analysis investigates the effect of variation in the model's parameters on the model's output by entering evidence and assessing its weight with respect to the output somehow. The idea of this procedure relies on temporary instantiation of evidence, or query, node followed by registration of the change in the probability distribution of the target variable. The total change in the probability distribution of the target variable summed with respect to all states of the evidence variable can be seen as the size of the influence of the query node on the target variable. Example measures of this type include measures like *value of information* [Pearl, 1988], or *weight of evidence* [Madigan *et al.*, 1997]. The most often used scores are Shannon's mutual information and quadratic score [Pearl, 1988].

The mutual information $I(T, X)$, also referred to as the entropy reduction between two variables: target variable T, and other variable X, is defined as the expected reduction in entropy of T (measured in bits) due to a finding at X, and can be calculated with the following expression:

$$I(T,X) = -\sum_{x}\sum_{t} P(t,x)\log_2 \frac{P(t,x)}{P(t)P(x)}, \tag{6.1}$$

where the summations are taken with respect to all possible instantiations (states) t, and x of T and X, respectively. The mutual information is thus measured in bits and takes a value from the range $[0, H(T)]$, where $I(T, X) = 0$ if T is independent of X, and $H(T)$ is the entropy of T before any new findings are entered.

For illustrative purposes let us consider the Bayesian network in Figure 6.2.1 that models the joint probability distribution over the set X of random variables X_i, $X = \{X1, X2, \ldots, X7\}$.

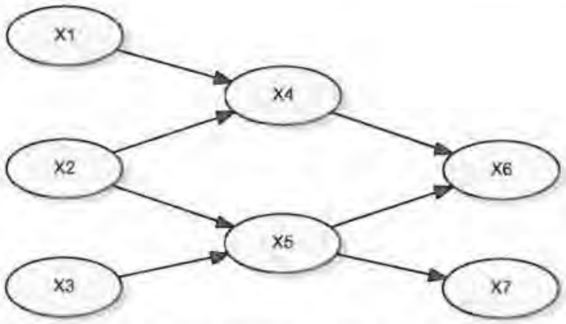


Figure 6.2.1. The Bayesian network structure used for the illustration.

Let us assume that each node X_i in the above model represent binary variables that can take values in the set {0, 1}. Furthermore, let us assume that the direct dependencies in this network are parameterised numerically with probabilities shown in Table 6.2.1.

Node		Numeric	Symbolic
X1	$p(X1=0)$	0.2	θ_{10}
	$p(X1=1)$	0.8	θ_{11}
X2	$p(X2=0)$	0.1	θ_{20}
	$p(X2=1)$	0.9	θ_{21}
X3	$p(X3=0)$	0.7	θ_{30}
	$p(X3=1)$	0.3	θ_{31}
X4	$p(X4=0\|X1=0, X2=0)$	0.2	θ_{4000}
	$p(X4=1\|X1=0, X2=0)$	0.8	θ_{4100}
	$p(X4=0\|X1=0, X2=1)$	0.6	θ_{4001}
	$p(X4=1\|X1=0, X2=1)$	0.4	θ_{4101}
	$p(X4=0\|X1=1, X2=0)$	0.7	θ_{4010}
	$p(X4=1\|X1=1, X2=0)$	0.3	θ_{4110}
	$p(X4=0\|X1=1, X2=1)$	0.3	θ_{4011}
	$p(X4=1\|X1=1, X2=1)$	0.7	θ_{4111}
X5	$p(X5=0\|X2=0, X3=0)$	0.1	θ_{5000}
	$p(X5=1\|X2=0, X3=0)$	0.9	θ_{5100}
	$p(X5=0\|X2=0, X3=1)$	0.8	θ_{5001}
	$p(X5=1\|X2=0, X3=1)$	0.2	θ_{5101}
	$p(X5=0\|X2=1, X3=0)$	0.7	θ_{5010}
	$p(X5=1\|X2=1, X3=0)$	0.3	θ_{5110}
	$p(X5=0\|X2=1, X3=1)$	0.6	θ_{5011}
	$p(X5=1\|X2=1, X3=1)$	0.4	θ_{5111}
X6	$p(X6=0\|X4=0, X5=0)$	0.8	θ_{6000}
	$p(X6=1\|X4=0, X5=0)$	0.2	θ_{6100}
	$p(X6=0\|X4=0, X5=1)$	0.6	θ_{6001}
	$p(X6=1\|X4=0, X5=1)$	0.4	θ_{6101}

	$p(X6=0 \mid X4=1, X5=0)$	0.5	θ_{6010}
	$p(X6=1 \mid X4=1, X5=0)$	0.5	θ_{6110}
	$p(X6=0 \mid X4=1, X5=1)$	0.6	θ_{6011}
	$p(X6=1 \mid X4=1, X5=1)$	0.4	θ_{6111}
X7	$p(X7=0 \mid X5=0)$	0.25	θ_{700}
	$p(X7=1 \mid X5=0)$	0.75	θ_{710}
	$p(X7=0 \mid X5=1)$	0.8	θ_{701}
	$p(X7=1 \mid X5=1)$	0.2	θ_{711}

Table 6.2.1 Conditional probability tables associated with the network in Figure 6.2.1.

The results of the empirical sensitivity of the node *X6* to findings at other nodes in the model presented in Figure 6.2.1 are shown in Table 6.2.2. From all the nodes, the most ability in reduction of the entropy of *X6* has the node *X4*.

Node	Mutual Info	Quadratic Score
X4	0.02569	0.0083082
X1	0.00120	0.0003939
X2	0.00010	0.0000340
X5	0.00009	0.0000308
X7	0.00003	0.0000087
X3	0.00002	0.0000050

Table 6.2.2. Sensitivity of the node *X6* due to a finding at another node.

The advantage of empirical sensitivity analysis lies in the fact that it provides the analyst with a concise summary measure that allows for consistent comparison of variables in terms of their influence on a target variable. On the other hand, it does not give any clue how much the output will change along with a slight variation in the particular parameter attached to the variable of interest.

In customer satisfaction studies the empirical sensitivity-based analysis is suitable if we want to simulate the effect of some marketing-mix actions on loyalty. For instance, if we used the model above to reason about the effect of a marketing-mix action, we could assume that *X1*, *X2*, and *X3* were some personality traits of a customer, *X4* and *X5* were some customer attitudes, and *X6* was the node representing success or failure of this action, we can find out which personality trait provides the most information as to the probability of success or failure of this marketing action. If we desire to reduce our uncertainty regarding the success of the marketing action in a specific country, on the basis of the results of the entropy based reduction, we could perform an examination of the prevalence of personality trait *X1* in the first place. Knowledge about the commonness of traits *X2* and *X3* in general does not contribute to our confidence in the success of the action. By the way, we note however that the chosen parameterisation of this example network yields that our knowledge of the specific state of variables *X1*, *X2*, or *X3* does not provide any substantial reduction in the uncertainty of *X6*.

6.2.2. Theoretical sensitivity analysis

The main focus in this study goes, however, to the other type of sensitivity analysis in Bayesian networks. The approach to sensitivity analysis referred to as *theoretical* aims at expressing the model's output as an algebraic function of the model's parameters. If the model's output in focus is the marginal probability $p(X_i=k)$ that the random variable X_i takes value k, then this approach tries to establish a function $f(p_m)$, such that

$$p(X_i=k) = f(p_m), \tag{6.2}$$

where p_m are model's parameters of interest. There can be of course one or more parameters of interest at a time.

In this context, the model's parameters denote some particular probabilities in the network – they can refer either to some particular entries in the conditional probability tables, or they can relate to prior marginal probabilities for nodes that have no parents.

It has been shown independently by Castillo *et al.* [1995, 1997] and by Coupe and van der Gaag [1997] that the sensitivity functions in Bayesian networks can be represented accurately with algebraic functions of a known form and unknown parameters. We will from now on refer to these unknown parameters as *coefficients* in order to distinguish them from the parameters-probabilities of interest. We will now address the work of Castillo *et al.* [1995] in more detail in order to cast more light on our application.

Let us start with the issue of symbolic propagation, as opposed to numeric propagation. Symbolic propagation leads to obtaining the marginal probabilities of interest that are expressed as functions of the parameters explicitly instead of real numbers. This kind of probabilistic inference requires often using of computer packages, which offer capabilities of symbolic computation, i.e. Maple or Mathematica. We will demonstrate the idea of symbolic propagation on our example Bayesian network model shown in Figure 6.2.1.

The joint probability distribution can be for this model in line with the chain rule of Bayesian networks (recall Formula 2.10) expressed as

$$\begin{aligned} p(X1,\dots,X7) = {} & p(X1)p(X2)p(X3)p(X4|X1, X2)p(X5|X2, X3) \\ & p(X6|X4, X5)p(X7|X5), \end{aligned} \tag{6.3}$$

Let us for the moment assume that all the entries in the conditional probability tables are treated as parameters. This means that all the nodes are symbolic. We will refer to these parameters as probabilities with the following symbols

$$\theta_{ij\pi} = p_i(X_i = j \mid \Pi_i = \pi_i), \tag{6.4}$$

where π_i are possible instantiations of the parents' set Π_i of the node X_i, and j = 0, 1, ..., r_i where r_i is the number of states of the node X_i. The first number in the subscript in $\theta_{ij\pi}$ refers to the node number, the second one refers to the state of the node, and the remaining numbers refer to the parents' instantiations. For simplicity, in case where a variable X_i does not have any parents, we will refer to its parameters with only two numbers in the subscript in θ_{ij}. All the symbolic parameters $\theta_{ij\pi}$ for our example network are presented in Table 6.2.1. Note that although there are 34 parameters in total, only 17 parameters are non-redundant, because the probabilities in each conditional distribution must sum up to unity.

Let us consult the prior marginal probabilities in the example model. Prior, also referred to unconditional, marginal probabilities refer to a situation in which there is no evidence in the network. If however some nodes are instantiated, then we will use the name posterior, or conditional marginal probabilities. We will also apply the term prior and posterior sensitivity analysis, respectively.

Some of the example prior marginal probabilities are shown in Table 6.2.3. The symbols shown in the column "Symbolic expression" refer to conditional probabilities from Table 6.2.1. To obtain these formulas we have made use of the factorisation of the joint probability distribution according to the Expression 6.3. To further simplify them we have taken account of the fact the probabilities in each conditional distribution must add up to one.

Marginal	Symbolic expression
$p(X1=1)$	θ_{11}
$p(X4=1)$	$\theta_{10}\theta_{20}\theta_{4100} + \theta_{10}\theta_{21}\theta_{4101} + \theta_{11}\theta_{20}\theta_{4110} + \theta_{11}\theta_{21}\theta_{4111} =$ $\theta_{10}\theta_{20}\theta_{4100} + \theta_{10}(1-\theta_{20})\theta_{4101} + (1-\theta_{10})\theta_{20}\theta_{4110} +$ $(1-\theta_{10})(1-\theta_{20})\theta_{4111} =$ $\theta_{10}\theta_{20}\theta_{4100} + \theta_{10}\theta_{4101} - \theta_{10}\theta_{20}\theta_{4101} + \theta_{20}\theta_{4110} - \theta_{10}\theta_{20}\theta_{4110} +$ θ_{4111}
$p(X7=1)$	$\theta_{710}(\theta_{20}\theta_{30}\theta_{5000} + \theta_{20}\theta_{31}\theta_{5001} + \theta_{21}\theta_{30}\theta_{5010} + \theta_{21}\theta_{31}\theta_{5011}) +$ $\theta_{711}(\theta_{20}\theta_{30}\theta_{5100} + \theta_{20}\theta_{31}\theta_{5101} + \theta_{21}\theta_{30}\theta_{5110} + \theta_{21}\theta_{31}\theta_{5111}) =$ $\theta_{20}\theta_{30}\theta_{5000}\theta_{710} + \theta_{20}\theta_{5001}\theta_{710} - \theta_{20}\theta_{30}\theta_{5001}\theta_{710} + \theta_{30}\theta_{5010}\theta_{710} -$ $\theta_{20}\theta_{30}\theta_{5010}\theta_{710} + \theta_{5011}\theta_{710} - \theta_{30}\theta_{5011}\theta_{710} - \theta_{20}\theta_{5011}\theta_{710} +$ $\theta_{20}\theta_{30}\theta_{5011}\theta_{710} - \theta_{20}\theta_{30}\theta_{5000}\theta_{711} - \theta_{20}\theta_{5001}\theta_{711} + \theta_{20}\theta_{30}\theta_{5001}\theta_{711} -$ $\theta_{30}\theta_{5010}\theta_{711} + \theta_{20}\theta_{30}\theta_{5010}\theta_{711} + \theta_{711} - \theta_{5011}\theta_{711} + \theta_{30}\theta_{5011}\theta_{711} +$ $\theta_{20}\theta_{5011}\theta_{711} - \theta_{20}\theta_{30}\theta_{5011}\theta_{711}$

Table 6.2.3 Symbolic expression of marginal probabilities.

For each expression of the marginal probability in this table we can see that it is stated as a polynomial function of the parameters. Formally, a polynomial in one variable (i.e., a univariate polynomial) with constant coefficients is given by the expression

$$a_n x^n + \ldots + a_2 x^2 + a_1 x + a_0, \tag{6.5}$$

whereas a polynomial in two variables (i.e., a bivariate polynomial) with constant coefficients is given by

$$a_{nm} x^n y^m + \ldots + a_{22} x^2 y^2 + a_{21} x^2 y + a_{12} x y^2 + a_{11} xy + a_{10} x + a_{01} y + a_0, \tag{6.6}$$

The highest power in a univariate polynomial is called its order, or sometimes its degree. A polynomial is an expression involving a sum of powers in one or more variables multiplied by coefficients.

From the form of expressions, we can notice that the parameters in every monomial are in the first degree (order). This is not accidental, as we have the following theorem [Castillo *et al.*, 1995]:

> *Theorem* 6.1. The prior marginal probability of any set of nodes is a polynomial in the parameters of degree less than or equal to the minimum of the number of parameters or nodes. However, it is a first-degree polynomial in each parameter.

Proof. According to the Chain rule of Bayesian networks (see Expression 4.2), the probability of any instantiation $(x_1, x_2, \ldots, x_n)$ is

$$\prod_{i=1}^{n} p(x_i \mid \pi_i), \tag{6.7}$$

that is, a product of n factors. Each factor is either $?_{ijp}$, if $x_i < r_i$ or $1-\sum_{j}^{r_i-1}\theta_{ij\pi}$ if $x_i = r_i$, that is, a parameter or a first degree polynomial in some parameters.

In addition, each parameter appears at most in one factor and dependent parameters, such as $?_{i0p}$ and $?_{i1p}$, do not appear in the same factor. Thus, we get a polynomial of degree less than or equal to the minimum of the number of parameters or nodes, which is first degree in each parameter.

Furthermore, the prior marginal probabilities of any node are polynomials in the parameters since the prior marginals are the sum of the probabilities of a subset of instantiations [Castillo *et al.*, 1995].

Let us now take a look at the case of posterior marginals, i.e., where some nodes are instantiated yielding the conditional marginals. The following theorem refers to the situation of posterior marginals.

> *Theorem* 6.2. The posterior marginal probability of any set of nodes Y, i.e., the conditional of the set Y given some evidence e, is a ratio of two polynomial functions of the parameters. Furthermore, the denominator polynomial is the same for all nodes.

Proof. The posterior marginal of any set of nodes $p(Y \mid e)$ is by definition given as

$$p(Y \mid e) = p(Y, e)/p(e)$$

so we have a rational function. On the basis of Theorem 1 we know that the probability of the evidence $p(e)$ is a polynomial function in parameters. Furthermore, the nominator is also a polynomial as it has been proved above.

In practice, however, we are interested in the marginal probabilities expressed only as a function of a few parameters of interest, whereas the other parameters take their numeric values. In this respect often, a distinction is made with regard to the number of parameters taken into account. One-way sensitivity analysis pertains to varying the value of just one parameter, whereas two-way sensitivity allows for examination of the strength of the influence of two parameters at a time.

It is important to note that the algebraic structures of the known form hold universally regardless of the particular dependency structure that the network encodes. Of course, not in every case the coefficients in sensitivity function are different from zero. In particular, in case any two variables under consideration are d-separated, and hence mutually independent, the variation of parameters assigned to one of them does not have any impact on the marginal probabilities of the other variable.

We discuss how the values of coefficients in the sensitivity functions can be found for one-way and two-way sensitivity analyses separately in the following paragraphs.

6.2.2.1. One-way sensitivity analysis

The theorems above guarantee thus that the marginal probabilities of any node is either a polynomial function or a ratio of polynomial functions. The general structure of a polynomial function is

$$\sum_{m_r \in M} c_r m_r \,, \tag{6.8}$$

where c_r is the numerical coefficient associated with the monomial m_r.

In accordance with Theorem 1, the target marginal probability of interest $p(X_i{=}k)$ is a polynomial function in the parameter of the first order and thus can be expressed using a linear function $p(X_i{=}k)(p_m)$ as

$$p(X_i{=}k)(p_m) = a_{ik} + b_{ik}p_m, \tag{6.9}$$

where p_m is the parameter-probability of interest whose influence on the target marginal $p(X_i{=}k)$ we want to exercise, and a_{ik} and b_{ik} are two meta-parameters.

An interesting thing is that we can establish the values of the coefficients using the numeric, standard propagation in Bayesian networks. There is thus no need of performing any symbolic computations with software capable of doing this, or determining the symbolic expressions as in Table 6.2.3. The easiest way to obtain the values parameters of the sensitivity functions using the numeric propagation methods is by solving a system of two linear equations for instance using the canonical components method. This can be presented with the matrix algebra style as

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a_{ik} \\ b_{ik} \end{pmatrix} = \begin{pmatrix} p(X_i = k)(0) \\ p(X_i = k)(1) \end{pmatrix}, \tag{6.10}$$

where $p(X_i{=}k)(0)$ is the value of the marginal probability $p(X_i{=}k)$ given the value of the parameter p_m is zero, and similarly $p(X_i{=}k)(1)$ is the value of the marginal probability $p(X_i{=}k)$ given the value of the parameter p_m is one.

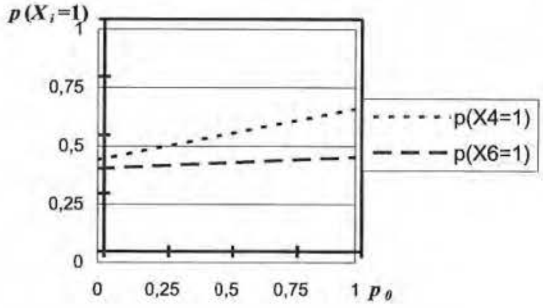
Returning to our example, let us take for the sake of illustration two parameters in focus, i.e. $p_0{=}p(X1{=}1)$, and $p_1{=}p(X5{=}1|X2{=}0, X3{=}1)$, and we treat the other remaining probabilities as constant by taking their numerical values shown in Table 6.2.1. The table below shows the one-way sensitivity functions of the node marginals in terms of each of both parameters separately.

	p_0	p_1
$p(X1=1)$	p_0	0.8
$p(X2=1)$	0.9	0.9
$p(X3=1)$	0.3	0.3
$p(X4=1)$	$0.44 + 0.22p_0$	0.616
$p(X5=1)$	0.366	$0.36 + 0.03p_1$
$p(X6=1)$	$0.40262 + 0.04962p_0$	$0.392216 + 0.0024p_1$
$p(X7=1)$	0.5487	$0.552 - 0.0165p_1$

Table 6.2.4 Marginal probabilities expressed in terms of parameters – one-way sensitivity analysis.

We can easily portray one-way sensitivity functions graphically using simple charts. For instance, charts in Figure 6.2.2 show the graphical representation of the sensitivity functions listed in Table 6.2.4.

a)



b)

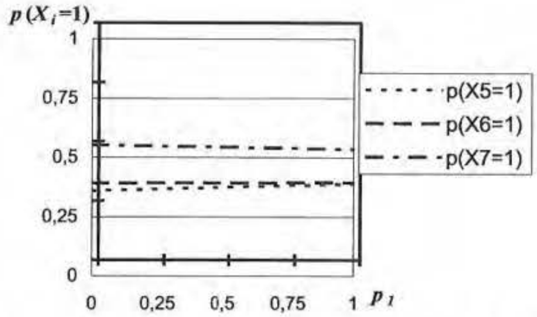


Figure 6.2.2 Graphical depiction of one-way sensitivity functions.

In Figure 6.2.2a) we present the sensitivities of marginals $p(X4=1)$ and $p(X6=1)$ as functions of the parameter p_0, and in 6.2.2b) the sensitivities of $p(X5=1)$, $p(X6=1)$ and $p(X7=1)$ as functions of the parameter p_1. In the charts above, the horizontal axis relates to the value of the parameter, whereas the vertical axis corresponds to the probability of the target marginal probability of interest. From the graphs in Fig. 6.2.2 we can read the lower and upper bounds of the marginal probabilities, which can be a valuable indication. We can moreover observe that some marginals in our example model are much sensitive to the parameters, whereas others hardly reveal any sensitivity.

6.2.2.2. Two-way sensitivity analysis

In line with the previous discussion, the two-way prior sensitivity function in a Bayesian network has the following algebraic structure:

$$p(X_i=k) = a_{ik} + b_{ik}p_{m0} + c_{ik}p_{m1} + d_{ik}p_{m0}p_{m1}, \tag{6.11}$$

where p_{m0} and p_{m1} are some parameters-probability of interest whose influence on the target marginal $p(X_i=k)$ we want to exercise, and a_{ik}, b_{ik} c_{ik} and d_{ik} are the coefficients to be determined.

Similarly to one-way sensitivity, we can determine the value of coefficients by solving a system of linear equations using the canonical components method. This can be presented with the matrix algebra style as

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a_{ik} \\ c_{ik} \\ b_{ik} \\ d_{ik} \end{pmatrix} = \begin{pmatrix} p(X_i = k)(0,0) \\ p(X_i = k)(0,1) \\ p(X_i = k)(1,0) \\ p(X_i = k)(1,1) \end{pmatrix}, \tag{6.12}$$

where, for example, $p(X_i=k)(0,0)$ is the value of the marginal probability $p(X_i=k)$ obtained when the value of both parameter p_{m0} and p_{m1} is zero.

For illustrative purposes, we will carry out the two-way sensitivity analysis for our example. The only node that is sensitive to both parameters turns out to be node $X6$. The functional form of its sensitivity is shown in Table 6.2.5.

Marginal	Functional form
$p(X6=1)$	$0.35324 + 0.04872p_0 - 0.0012p_1 + 0.0045p_0p_1$

Table 6.2.5 Prior marginal probability $p(X6=1)$ expressed as a function of two parameters.

Let us picture the dependency of marginal probability $p(X6=1)$ with a chart. Figure 6.2.3 represents this dependency.

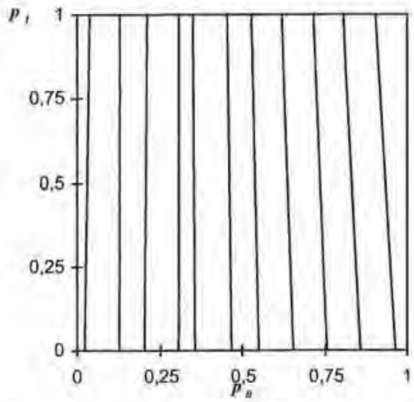


Figure 6.2.3 Two way sensitivity analysis.

When identifying interaction effects, the main focus goes to the sign and size of the coefficient d_{ik} (see Expression 6.11). Positive values of this parameter stand for positive synergy, whereas negative values stand for negative interaction effects. Values close to zero may indicate a lack of interaction effects between product/service dimensions. Additional insight might be achieved by studying interaction effects among a set of three and even more parameters at a time. Higher-order sensitivity analyses are however less often used in practice due to complexity and cumbersome interpretation of their results.

6.2.2.3. Conditional sensitivity functions

As a result of Theorem 2 the algebraic form of the posterior sensitivity function is a ratio of two polynomial functions. Consequently, in the case of one-way sensitivity analysis the following form holds:

$$p(X_i = k \mid e)(p_m) = \frac{a_{ik} + b_{ik} p_m}{c_{ik} + d_{ik} p_m}, \tag{6.13}$$

where $p(X_i = k| \; e)(p_m)$ is the target marginal probability of interest, e is evidence, p_m is the parameter of interest, and a_{ik}, b_{ik}, c_{ik} , and d_{ik} are the coefficients (meta-parameters).

To calculate the coefficients in the one-way sensitivity analysis, we can first determine the actual values of the coefficients c_{ik} and d_{ik} in the denominator in a way similar as in the case of prior sensitivity functions, i.e. by finding the sensitivity of $p(e)$ as a function of p_m. Then, we can determine the values of the coefficients a_{ik} and b_{ik}. For instance, using the canonical components method, first we set the parameter p_m to zero, we update the probabilities in the network, we note the value of $p(X_i = k| \; e)$ and find $a_{ik} = cp(X_i = k| \; e)$. Next, we set the parameter p_m to one, again update the model, we note the new value of $p(X_i = k| \; e)$ and find $b_{ik} = (c_{ik}+d_{ik})p(X_i = k| \; e)-a_{ik}$.

Given evidence e={$X4$ = 1}, the one-way sensitivity of some selected variables for our example Bayesian network model is shown in Table 6.2.6.

Marginal	Functional form
$p(X5{=}1 \mid X4{=}1)$	$\dfrac{0.174+0.0546p_0}{0.44+0.22p_0}$

Table 6.2.6 Marginal probabilities expressed with parameters.

As regards the posterior two-way sensitivity analysis function, the following algebraic form holds:

$$p(X_i = k \mid e)(p_{m0}, p_{m1}) = \frac{a_{ik} + b_{ik} p_{m0} + c_{ik} p_{m0} + d_{ik} p_{m0} p_{m1}}{e_{ik} + f_{ik} p_{m0} + g_{ik} p_{m1} + h_{ik} p_{m0} p_{m1}}, \tag{6.14}$$

where $p(X_i = k| \; e)(p_{m0}, p_{m1})$ is the target marginal probability of interest, e is evidence, p_{m0} and p_{m1} are parameters, and a_{ik}, b_{ik}, c_{ik} , d_{ik}, e_{ik}, f_{ik}, g_{ik}, and h_{ik} are the coefficients (meta-parameters). The calculation of these values of these coefficients can be achieved accordingly with the method suggested earlier.

Table 6.2.7 shows the sensitivity of $p(X5{=}1)$ given the evidence $e{=}\{X6 = 1\}$.

Marginal	Functional form
$p(X5{=}1 \mid X6{=}1)$	$\dfrac{0.144+0.012p_1}{0.35324+0.04872p_0-0.0012p_1+0.0045p_0p_1}$

Table 6.2.7 Marginal probabilities expressed with parameters.

In the sequel, we apply the theoretical approach to sensitivity analysis in Bayesian networks to acquire analytical knowledge from the model.

6.3. Classification of features

To complete the analysis of feature importance, we should define a relevant feature classification scheme. There exists a number of studies suggesting various feature classification schemes. For instance, in [Levitt, 1983] a four-ring conceptualisation of a product/service is suggested as a unitary concept, according to which the most inner ring represents the generic product – a must. The next ring defines the expected product, comprising dimensions acting as satisfiers/dissatisfiers. The augmented or enhanced product surrounds the expected product attributes, and acts as delights to a customer. Most valuable insights to a marketer are delivered however with the outermost ring that determines the potential product, i.e. the product that should contribute most to the company's success in the future.

In this study, we adopt the classification of attributes from Vanhoof and Swinnen [1996], in which the authors proposed the following four categories of service attributes (or dimensions): 1) satisfier/dissatisfier, 2) exciter, 3) basic, and 4) non-relevant. As a *satisfier/dissatisfier*, we will regard a dimension that affects satisfaction in its continuum, i.e. both its high and low levels, thus driving high levels of satisfaction when performed well and enforcing dissatisfaction when their perception falls below expectations. Moderate or large influence on high overall satisfaction, and insignificant effect on dissatisfaction characterizes features that can be defined as an *exciter*. Exciters are drivers of satisfaction as well, but they do not influence dissatisfaction if their performance is low. If, in turn, high overall satisfaction is not affected by high feature perception, and if at the same time dissatisfaction is likely to intensify when this perception is low, the feature can be viewed as a *basic* product dimension delivering elementary user's requirements. As the feature performance does not make any changes in perception of overall (dis)satisfaction, it can be interpreted as *non-relevant*.

We will now explain the categories just mentioned within the framework of sensitivity analysis in Bayesian networks. In line with the theoretical one-way prior sensitivity analysis, each level of overall satisfaction can be captured with a linear function. Given that the overall satisfaction is modelled with three distinct levels, e.g., low, medium, and high, the probability of overall satisfaction $p(Sat)$

at each of these levels can be described as a function of the satisfaction with a service dimension X with the following form:

$$\begin{gathered} p(Sat=\text{'low'}) = a_l + b_l p(X=\text{'low'}), \\ p(Sat=\text{'medium'}) = a_m + b_m p(X=\text{'medium'}), \\ p(Sat=\text{'high'}) = a_h + b_h p(X=\text{'high'}), \end{gathered} \tag{6.15}$$

where $p(X=\text{'high'})$ is probability that the satisfaction with the service dimension is high, $p(X=\text{'medium'})$ is probability that the satisfaction with the service dimension is medium, $p(X=\text{'low'})$ is probability that the satisfaction with the service dimension is low, and the parameters a_l, a_m, and a_h amount to the probability of low, medium, or high satisfaction given the probability of respective level of satisfaction with the service dimension is zero. The linear coefficients b_l, b_m, and b_h can be interpreted as a measure of how relevant, or important, the service dimension is with regard to satisfaction at a specific level. Of course, the higher the absolute value of these parameters, the more influential the item is with regard to (dis)satisfaction.

The categories can be defined according to the values of parameters b_h and b_l in the functions above. These categories are shown in Table 6.3.1. Whether the influence is low, moderate, or large can be determined by looking at the absolute values of parameters b_h and b_l. We assume that high satisfaction with a service dimension has a negative (non-increasing) effect on low overall satisfaction, and a positive (non-decreasing) impact on high overall satisfaction.

b_l \ b_h	Low	Moderate/Large
Low	Non-relevant	Exciter
Moderate/Large	Basic	Satisfier/Dissatisfier

Table 6.3.1. Categories of service elements with respect to values of parameters b_l and b_h in sensitivity functions.

As defined by Table 6.3.1, any dimension for which the sensitivity functions give low values of the coefficients b_h and b_l can be regarded as non-relevant to overall satisfaction. If the coefficient b_h is low and b_l is moderate, or high, then this dimension can be viewed as basic. Conversely, if b_h is high and b_l is low, then we deal with an exciter. At last, if both coefficient values are high, then this dimension is a satisfier/dissatisfier. Of course the assessment whether the values fall into low, or high range is subjective, and can vary from study to study.

6.4. Importance-performance analysis

Having identified importance of dimensions, the next step in the customer satisfaction measurement study is to determine its actual performance. Analysis of dimension's performance along with its importance can be combined to form conclusions, which can help to focus company resources on priorities for improvement with the purpose of fostering customer satisfaction [Hill and Alexander, 2000]. With this end in view, we can relate the score of satisfaction with dimensions with the score for general satisfaction. A method with which relative performance v can be calculated is with the following formula:

$$\upsilon := \frac{p(D=\text{"high"}) - p(OS=\text{"high"})}{p(OS=\text{"high"})}, \tag{6.16}$$

where $p(D=\text{"high"})$ is the probability of high dimension's performance rating, and $p(OS=\text{"high"})$ is the probability of high general performance, i.e. overall satisfaction, rating.

Value v lower than -0.1 is considered as low, and a value higher than 0.1 is considered high; value v in between can be regarded as a moderate value. The calculation in Formula 6.16 takes into account only probabilities of high performance ratings. However, overall satisfaction is defined as a variable with low, moderate, and high rating, whereas service dimensions are binary.

The marketing literature is relatively rich in the numerous classifications of features with regard to relation they have on product and company success. For instance, Ortinau *et al.* [1989] propose four categories: "concentrate here", "keep up the good work", "low priority", and "possible overkill". We adapt the following similar categorization from Vanhoof and Swinnen [1996] presented in Table 6.4.1.

Performance	Relevance	Category
Low, moderate	Non-relevant	Low priority
Low, moderate	Satisfier/Dissatisfier, Basic	Action needed
Low, moderate	Exciter	Opportunities
High	Satisfier/Dissatisfier, Exciter	Strengths
High	Basic	Take care
High	Non-relevant	Possible overkill

Table 6.4.1. Categorization of product/service dimensions with respect to their priorities for company management.

The category *low priority* can be assigned to those antecedents of overall satisfaction that exhibit low performance and are non-relevant to overall satisfaction. *Action needed* pertains to features not performing well, but acting as dissatisfier or basic dimension, thus requiring more company resources. Customers' delight and high levels of satisfaction can be fostered by focusing on *opportunities*. *Company strengths* are those attributes that have high performance and act as either satisfier/dissatisfier or exciter. *Take care* attributes are having a high strong influence on dissatisfaction and high performance, thus their performance should be continually supported. Non-relevant attributes having high performance are likely to be excessively paid attention to and can be classified into *possible overkill* category.

6.5. Data issues

The study is situated in the business-to-customer phone service industry. As follows from the recent customer survey [Mobius, 2002], customer satisfaction should be an essential objective for phone operators.

6.5.1. Data collection

The collection of data used in this study has been conducted by a marketing research agency for a telecom company (Tritone Telecom) operating a fixed phone line in the Netherlands for the purpose of a customer satisfaction study. Potential respondents were chosen from among the company clients and asked by phone to participate in a customer satisfaction study. Originally, 523 clients responded positively to the survey and took part in it.

6.5.2. Questionnaire

The questionnaire was aimed at collection of customer responses with respect to overall customer satisfaction, loyalty, and satisfaction with various aspects of the service, e.g., sales force, connections, customer service, tariffs, and billing. Overall satisfaction has been measured with one item, whereas satisfaction with the respective dimensions has been captured in terms of specific service features relating to those dimensions on a 5-point Likert-type scale anchored with "very satisfied" and "very dissatisfied".

Construct	Items
Overall Satisfaction (S)	1.How satisfied are you in general with your telecom operator?
Customer Service (CS)	1.How satisfied are you with the reaction time of the customer service at ... in case of problems or enquiries? 2.How satisfied are you with the speed with which a problem was solved? 3.How satisfied are you with the quality of the reply from the customer service? 4.How satisfied are you with the friendliness of the customer service? 5.How satisfied are you with the contacting person?
Contact with Customer Service	1.Have you contacted the Customer Service by phone during the past 6 months?
Tariffs (T)	1.How satisfied are you with the price/quality ratio with respect to national calls? 2.How satisfied are you with the price/quality ratio with respect to calls to mobile phone? 3.How satisfied are you with the price/quality ratio with respect to grensregio calls? 4.How satisfied are you with the price/quality ratio with respect to International calls?
Bills	1. Do you receive bills personally?
Billing (B)	1.How satisfied are you with the clarity of bills? 2.How satisfied are you with the amount of information on the bill?

Table 6.5.1 Operationalization of the constructs.

As regards satisfaction with customer service and billing, the questionnaire elaborated whether it is reasonable to ask the respondent about his satisfaction with these dimensions. In cases when the respondent did not contact the customer service, and did not receive bills personally it makes no use to ask about his satisfaction in these regards.

Tariffs can be viewed as a choice attribute, thus not relevant to formation of satisfaction judgments. However, we have decided to include it in the set of satisfaction attributes, because often the customer is not capable of choosing the most convenient rate scheme a priori, and the resultant disconfirmation of expectations in this respect may influence satisfaction judgments on tariffs.

Satisfaction with tariffs was captured based on the responses on the customers' satisfaction with four types of telephone connection tariffs: international, national, regional, as well as tariffs on connections to mobile phones. Based on perceptions of satisfaction with reaction time, service time, and quality of assistance another variable reflecting satisfaction with customer service was derived. Responses on amount of information and clarity were used to create customer's evaluation of billing service.

6.5.3. Data manipulation

First, all the responses for all the features as well as for overall satisfaction have been aggregated from five to three categories in order to facilitate interpretation and parameter learning. Levels of "very dissatisfied", "dissatisfied", and "neither satisfied nor dissatisfied" due to their low response frequency have been grouped together and assigned one value "low satisfaction". The scores of "satisfied" and "very satisfied" have obtained the meaning of moderate and high satisfaction, respectively.

Because the satisfaction scorings at a service dimension level, i.e. overall satisfaction with customer service, tariffs, and billing have not been operationalized by the questionnaire, in the next step three additional variables have been created to represent overall judgments of satisfaction with these dimensions. Satisfaction with billing service, satisfaction with tariffs and satisfaction customer service were obtained by clustering of respondents using *k*-means algorithm. We have decided that the satisfaction at the dimension level be only binary, instead of ternary. The reason for it was that it is probably easier to figure out and to interpret the results. Each construct obtained by this procedure had the centres reflecting the categories of low and highly satisfied customers. From the original sample we have removed 95 cases due to having more than 50% of missing values, what resulted in the final sample of 428 cases.

6.5.4. Model specification

In the next step of the analysis, we have constructed a small Bayesian network for the scenario under consideration consisting of four nodes. In accordance with the presupposed domain knowledge described in the previous paragraphs, we supposed that tariffs, billing and customer service are causes of the overall satisfaction the phone services of the supplier. Therefore, in our model, overall satisfaction is a child node of three nodes: satisfaction with customer service, billing and tariffs. Doing so we intentionally allow for relationships between satisfaction with the dimensions and overall satisfaction; we validate this model empirically later. The three causes are furthermore marginally independent, but they become dependent once value of overall satisfaction is fixed. For clarification purposes, we should note that the constructed BN model enables investigating all the three dimensions in one model, even though the data for some cases on Billing and Customer Service are structurally missing. Therefore, the Bayesian network modelling allows for the full use of all available data in the sense that it accounts for both satisfaction of respondents who contacted company's customer service for assistance, and thus provided answers on this subject, and of those customers who did not have a need to contact it and to whom questions about customer service are not applicable. The model structure is shown in Figure 6.5.1.

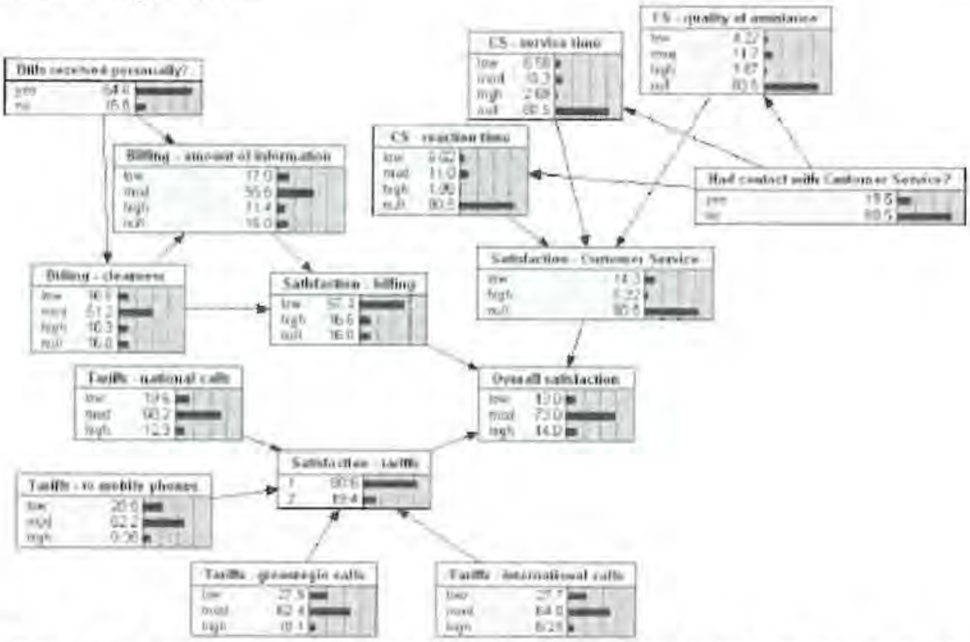


Figure 6.5.1. A Bayesian network structure and marginal probabilities for the data under study. The state labelled 'null' refers to users for whom the service is not applicable.

The numerical strengths of the dependencies, i.e., the conditional probabilities in the model, have been estimated based on maximum likelihood approach using the EM procedure to deal with missing data [Lauritzen, 1995].

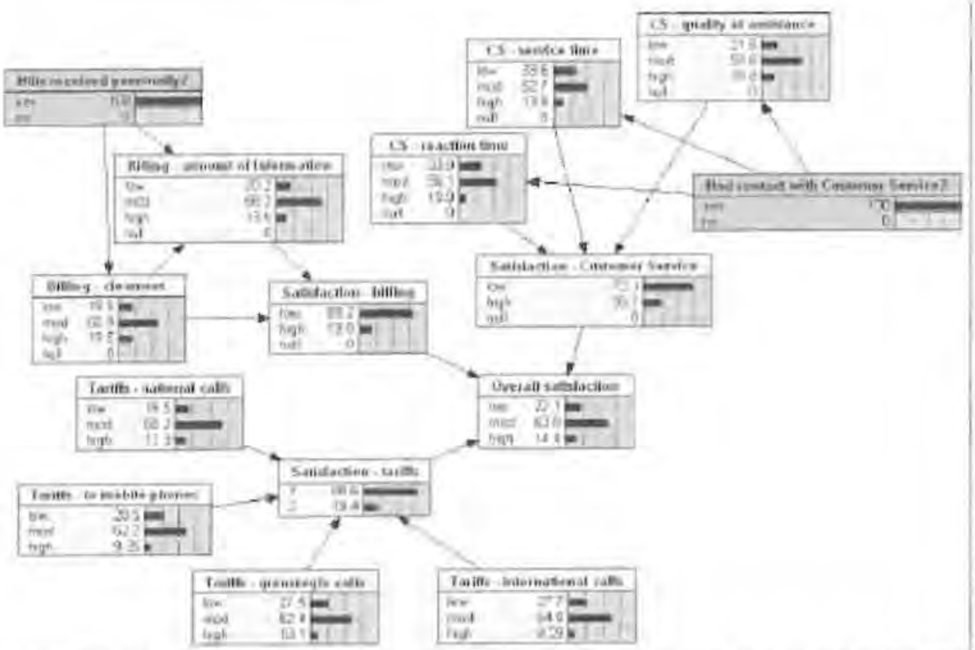


Figure 6.5.2 Marginal probabilities for repondents who contacted customer service and receive bills personally.

6.6. Results

6.6.1. Empirical validation

The Bayesian network model of any system can be viewed as a decision model and thus validated against empirical data by using it as a classifying system, in which the value of each variable for each case in the test set is predicted based on values of other observed variables. The goodness of fit of such a system is assessed by measuring its standard predictive accuracy, i.e., percentage of cases classified correctly, or alternatively using quadratic loss (Brier) score.

A good practice is to treat each node sequentially as a decision class, and use the model to predict the label of each case using 10-fold cross-validation. The method selects each time randomly 10% of the cases, uses the remaining cases to learn about the model's parameters, and finally applies the model to classify the case based on values of other variables. This procedure is repeated 10 times for each node.

Since each classification decision in the above process is probabilistic in nature, its outcome depends heavily on the probability distribution for states of the target node. To account for the uncertainty, and to overcome the deficiency of standard measure of predictive accuracy in this respect, another measure, known as Brier score, for assessing probabilistic decision systems was introduced [Panovsky and Brier, 1968]. The intuitive idea behind the Brier score is that in case the posterior probability of a specific category of overall satisfaction is remarkably higher than for the other categories and the prediction is correct, then the quality of such a forecast is better as if the distribution of categories was more resembling uniform distribution [Gaag and Renooij, 2001].

	Accuracy	Brier Score
Tariffs	99.7%	0.01019
Billing	100%	0.0001172
Customer Service	99.0%	0.01342
Overall Satisfaction	75.8%	0.3596

Table 6.6.1. Results of the empirical validation of the model under study.

We have applied the approach to validation as outlined above. On the whole, i.e., averaging the results of prediction of all the nodes in the model, predictive accuracy of 84% was obtained, and for service dimensions the performance was about 99%. Such a high result is a consequence of the fact that service dimensions are to an extent dummy nodes created by clustering. For overall satisfaction a score of 75.8% correctly classified cases was achieved, whereas the Brier score amounted to 0.3596.

To objectively interpret these outcomes, we should compare them with two other less informed classification models [Gaag and Renooij, 2001]. The first classifier based on the uniform probability distribution of overall satisfaction categories for each case gives the accuracy of 73% and the Brier score of 0.37. For the second model encoding marginal prior probability distribution of satisfaction, accuracy of 73,14% and a Brier score of 0.429 is obtained. Therefore we can conclude that our model is well calibrated and can be utilized in the feature performance analysis for this study.

Other alternative validation methods are usually based either on Bayesian scores for a network structure, or on properties of (un)conditional independencies among vertices in a network. We have found that the structure of the model in focus was supported by the assertions of (un)conditional independence properties determined from empirical data by the PC algorithm [e.g., Spirtes *et al.*, 2001].

6.6.2. Marginal probabilities

In Figure 6.5.2, we present the marginal probabilities in the model given the respondent has contacted customer service and receives bills personally. In the sequel, we will perform the analysis for this group of respondents, including thus all the three service dimensions.

6.6.3. Classification of service dimensions

Since in our study, the overall satisfaction can take three different levels, i.e., low, medium, and high, we can express the probability of each of these levels as a separate function. The most important for our analysis is of course the influence of service dimensions on high and low overall satisfaction. The medium level of overall satisfaction can be seen as a buffer.

We will denote the probability of high satisfaction with customer service with p_{CS_high}, the probability of high satisfaction with tariffs with p_{T_high}, and the probability of high satisfaction with billing as p_{B_high}. Accordingly, the probability of low satisfaction with customer service with p_{CS_low}, the probability of low satisfaction with tariffs with p_{T_low}, and the probability of low satisfaction with billing as p_{B_low}. The probability of the medium level of overall satisfaction has been assessed as function of probabilities of high satisfaction with dimension. Table 6.6.2 contains the functional form of the sensitivity of each possible degree of overall satisfaction.

Marginal	Function
$p(S=\text{'low'})$	$0.137 + 0.113p_{CS\ low}$
	$0.145 + 0.093p_{B\ low}$
	$0.146 + 0.092p_{T\ low}$
$p(S=\text{'medium'})$	$0.670 - 0.130p_{CS\ high}$
	$0.688 - 0.267p_{B\ high}$
	$0.591 + 0.270p_{T\ high}$
$p(S=\text{'high'})$	$0.078 + 0.243p_{CS\ high}$
	$0.072 + 0.361p_{B\ high}$
	$0.169 - 0.134p_{T\ high}$

Table 6.6.2. Functional form of dependencies.

To gain better insight into the relations between the service dimensions and overall satisfaction, let us picture these functions graphically. In Figure 6.6.1, the probability of high satisfaction with a service dimension is shown on the horizontal axis, whereas the vertical axis is the probability of the relevant level of overall satisfaction. From these graphs we can read the boundaries between which specific levels of the overall satisfaction can vary as a result of feature performance. For instance, the probability of high overall satisfaction varies from 0.07 to 0.31 as a result of bad and good customer service, respectively. Also, on the basis of the observation that both dissatisfaction (Figure 6.6.1a) and high satisfaction (Figure 6.6.1c) are sensitive to changes in customer service performance ($|b_l| = 0.11$, $|b_h| = 0.24$), we conclude that customer service can be classified as satisfier/dissatisfier. Similarly, we can classify billing also to the same category ($|b_l| = 0.09$, $|b_h| = 0.36$), whereas tariffs, due to their positive impact on moderate satisfaction and negative impact on high satisfaction, warrant a closer look to arrive at the right conclusion. Nevertheless, billing quality has a more substantial impact on satisfaction than customer service has.

a) $p(S=\text{'low'})$

b) $p(S=\text{'mod'})$

c) $p(S=\text{'high'})$

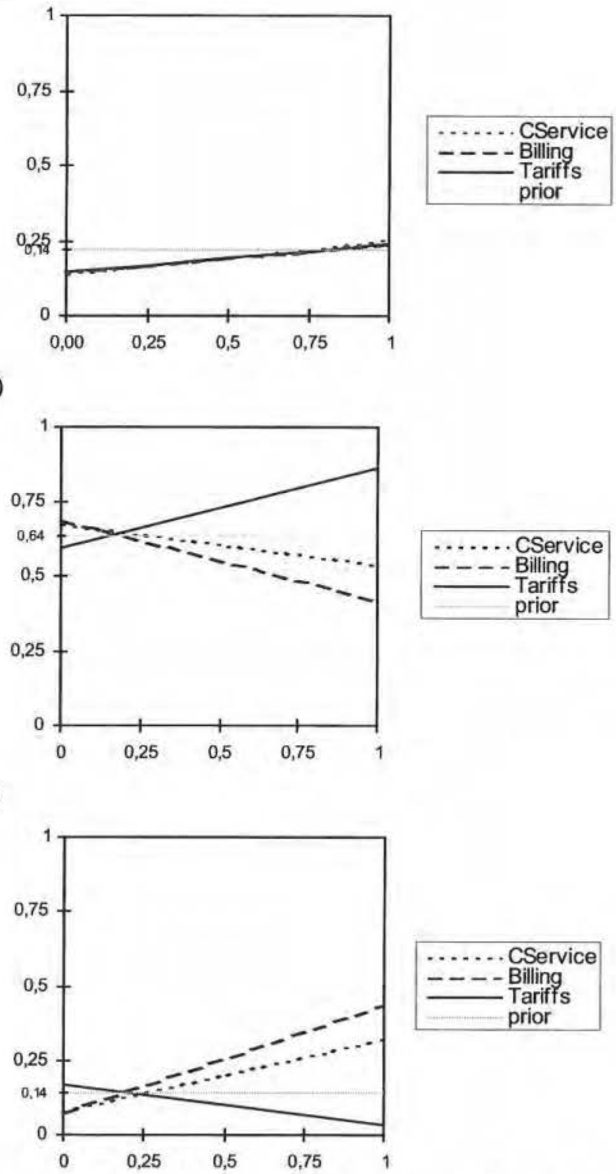


Figure 6.6.1. Impact of service elements on: a), low, b), moderate, and c), high levels of overall satisfaction, respectively. The grey lines represent prior probability of the respective level of satisfaction.

The charts show also the prior levels of given satisfaction, that is, the present level reflected by the data.

These graphs confirm the findings in [Mittal *et al.*, 1998] in that they show the diverse nature of the influence of satisfaction with a feature on overall service satisfaction: low levels of satisfaction are found hardly sensitive to dissatisfactory experiences with service dimensions, whereas high overall satisfaction shows in this respect an increased dependence.

6.6.4. Two-way sensitivity

It is likely that some potential determinants of overall satisfaction do not manifest an apparent influence when considered apart from other factors. It can however at the same time happen to be an important factor catalysing the impact of other service dimensions. Synergy effects that can be observed in this situation may be either positive or negative. Their existence can be traced by means of two- and multi-way sensitivity analysis.

Recall from Section 6.2.2 that the general form of two-way sensitivity can be expressed as

$$p(Z{=}z) = a + b\,p(X{=}x) + c\,p(Y = y) + d\,p(X{=}x)\,p(Y = y), \qquad (6.17)$$

$p(Z{=}z)$ is the target probability of interest that variable Z takes state z, $p(X{=}x)$ and $p(Y{=}y)$ are probabilities that the explaining variables X and Y take states x and y respectively, and a, b, c, and d are meta-parameters to be calculated by performing inference in the network. The coefficients of the sensitivity functions can also be used to classify the two-way interaction. Parameter a can be interpreted as a probability of high overall satisfaction, when neither dimension is satisfactory. Parameters b and c have a similar interpretation as in one-way sensitivity functions and can be used to determine whether one service element is dominant over another.

Translating this into importance-performance framework, the sensitivities at each level of general performance (satisfaction) can be different for the different target values, so we have to calculate the following sensitivity functions:

$$\begin{aligned} p(S{=}\text{low}) &= a_l + b_l p(X{=}\text{low}) + c_l p(Y{=}\text{low}) + d_h p(X{=}\text{low})p(Y = \text{low}),\\ p(S{=}\text{mod}) &= a_m + b_m p(X{=}\text{mod}) + c_m p(Y{=}\text{mod}) + d_m p(X{=}\text{mod})p(Y{=}\text{mod}), \qquad (6.18)\\ p(S{=}\text{high}) &= a_h + b_h p(X{=}\text{high}) + c_h p(Y{=}\text{high}) + d_h p(X{=}\text{high})p(Y{=}\text{high}), \end{aligned}$$

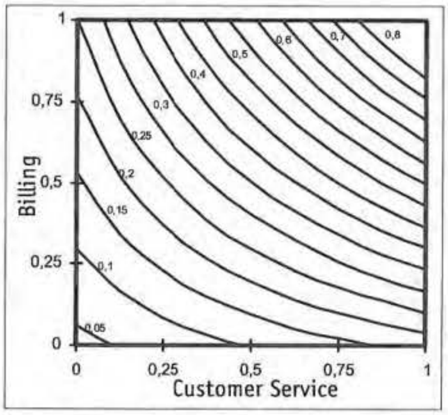
where X and Y stand for service dimensions, S refers to overall satisfaction, $p(X{=}\text{'high'})$ is the probability that the satisfaction with the service dimension X is high. The parameters a_l, a_m, and a_h amount to the probability of low, medium and high overall satisfaction, when the probability of satisfaction with respective service dimension equals zero.

The sensitivity functions in Expression 6.18 can also be represented graphically (see Figure 6.6.2). The graphs represent the sensitivity of high overall satisfaction judgments to variation in the perception of three service dimensions: customer service, billing quality and connection tariffs. Simultaneous variation of two probabilities resulting in the same probability of high overall satisfaction is represented by the contour lines, and the numbers attached to the lines stand for the probability level.

In Figure 6.6.2a), for instance, the probability that a customer is satisfied with the customer service is shown on the X-axis and with the billing service on the Y-axis. The upper rightmost contour line denotes that all the combinations of (high) probabilities with feature performance located on this line result in the high, as of 80%, value of probability of high overall satisfaction. The lower leftmost line corresponds to the combination of rather high probabilities of dissatisfactory experience at each dimension level. In that case, the probability of high overall satisfaction amounts to 3%. The numerical properties of the sensitivity function communicate that this variation ranges from 3% up to 92%. The slope of the lines suggests further that in the low ranges of customer service performance, overall satisfaction is much less sensitive to changes in perception of billing than to customer service. However, in the higher ranges, this relation reverses, and on the whole, billing has more influence than customer service. This is evidenced in the parameters b=0.13 and c=0.21. Finally, because the lines at the higher ranges of explaining probabilities get closer to each other and the resulting probability gets higher we can observe a joint interaction effect. This is confirmed by the value of parameter d_h=0.55.

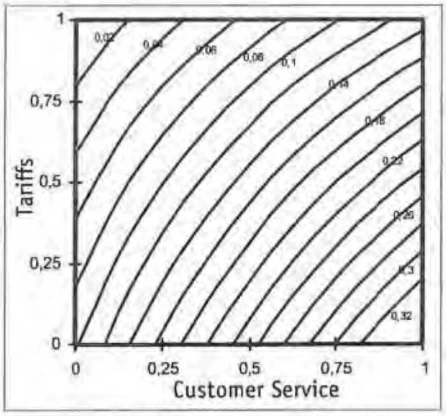
We can thus infer that the better the perception of both service dimensions, the more positive the satisfaction judgments. Figure 6.6.2b) shows that the probability of high overall satisfaction as a result of customer service and tariffs can vary from about 1% to 35%. The lowest probability is achieved as a result of a dissatisfactory experience with customer service and very high chance of satisfaction with the tariffs. This situation shows a strong negative synergy (d_h=-0.14). In Figure 6.6.2c) the contour lines are drawn nearly in parallel every 5% and vary from 2% to 47% implying high and constant sensitivity of high satisfaction to varying performance of billing and tariffs. By comparing the graphs we can infer again that the most important dimension is billing, which explains most variation in overall satisfaction when compared to other dimensions.

a)



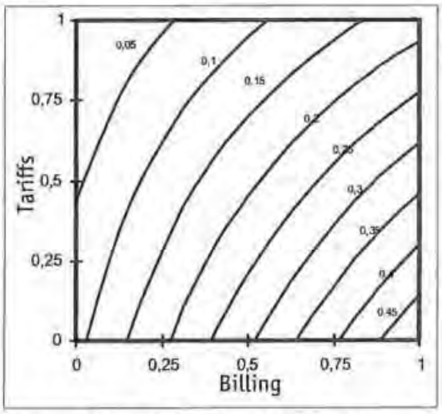
$$p(OS=\text{'high'}) = 0.03 + 0.13\, p(CS=\text{'high'}) + 0.21\, p(B=\text{'high'}) + \\ + 0.55\, p(CS=\text{'high'})\, p(B=\text{'high'})$$

b)



$$p(OS\text{='high'}) = 0.09 + 0.27\, p(CS\text{='high'}) - 0.09\, p(T\text{='high'}) - 0.14\, p(CS\text{='high'})\, p(T\text{='high'})$$

c)



$$p(OS\text{='high'}) = 0.09 + 0.4\, p(B\text{='high'}) - 0.89\, p(T\text{='high'}) - 0.22\, p(B\text{='high'})\, p(T\text{='high'})$$

Figure 6.6.2. Interaction effects between a), customer service (X-axis) and billing (Y-axis), b) customer service and tariffs, c) billing and tariffs. The contour lines correspond to combinations of probability of the satisfactory service dimension that result in the same probability of high overall satisfaction.

6.6.5. Performance

In order to calculate the performance of the service dimensions, we compared their marginal probability distributions with the one for overall satisfaction applying transformation 6.16. Since all the three performance factors fall below -0.1 their performance can be classified as low.

From the importance-performance analysis it follows that the company should undertake some actions to improve the performance of the considered service aspects billing being the first priority, and customer service being the second.

Further insight regarding phone tariffs is required to formulate a relevant marketing policy in this respect.

6.6.6. Simulations

The analysis presented hitherto concerns exclusively customers who manifested having got in touch with customer service and thus it shows how overall satisfaction can be affected by (un)satisfactory encounter with customer service. However, not all clients engage in contact with customer service, even when they seek advice or experience problems during the service delivery. A marketing manager may be therefore interested to know what is the influence of good customer service on overall satisfaction by contrasting satisfaction responses of both groups with each other. If it turns out that customers manifesting satisfaction tend to be much more satisfied overall than customers who do not seek assistance at customer service, it can be a good initiative to encourage clients even stronger to engage in contact with company in case of any questions or problems. Marketers know a similar regularity that clients who experience problems during service delivery and complain about them to the company, and subsequently receive redress, tend to be loyal and engage in word-of-mouth recommendation behaviour. The Bayesian network methodology accounts also for satisfaction responses of clients who did not contact customer service. The significance of quality customer service can be assessed by comparison of overall satisfaction responses by two heterogeneous groups of customers: those who probably experienced a problem with their telephone connections and subsequently contacted customer service for assistance with those who did not contact customer service.

Group of customers	Overall Satisfaction		
	low	moderate	high
Overall	0.13	0.74	0.13
Customers who: - did not contacted CS - did contacted CS	 0.10 0.23	 0.76 0.62	 0.14 0.15

Table 6.6.3. Probability values for overall satisfaction with respect to customers who have contacted Customer Service and who have not.

The probability distributions with respect to the two groups are shown in Table 6.6.3. If we do not know whether a particular consumer contacted customer service or not, we infer that she should be dissatisfied with probability 13%, moderately satisfied with probability 74%, and very satisfied with probability 13%. For the group of customers who did not contact customer service, the probabilities are very similar as for the entire dataset, i.e. 10% of them tend to be dissatisfied, 76% are moderately satisfied, and 14% very satisfied. Customers who did contact customer service display, on the whole, have different characteristics. As much as 23% of this group becomes dissatisfied, 62% are moderately satisfied, and 15% is very satisfied.

The significance of having good customer service becomes apparent if we delve into the responses by the group of clients who provided answers on this dimension. This can be read from Table 6.6.4.

Customers who contacted CS	Overall Satisfaction		
and are:	low	moderate	high
- dissatisfied	0.27	0.64	0.09
- satisfied	0.12	0.56	0.32

Table 6.6.4. Comparison of overall satisfaction levels for customers who perceive Customer Service as satisfactory and dissatisfactory.

It turns out that clients dissatisfied with customer service are not likely to be very satisfied in general, since this probability amounts only to 9%. On the contrary, they tend to be dissatisfied with probability 27%, and moderately satisfied overall with probability 64%. The customers, who experienced positive customer service resulting in satisfaction with this aspect, are very likely to perceive satisfaction. In total, 88% of them reported satisfaction, of which 32% can be regarded as very satisfied overall, and 56% tend to be moderately satisfied. At the same time only 12% of those clients seem to be dissatisfied in general.

As a result, we can conclude from this simulation that fulfilling expectations towards customer service encounter pays off very well, since it is more likely that clients become more satisfied than those who do not engage in contact with it. Furthermore, it turns out that the company should improve its customer service performance, because in general those customers who contacted CS are still more likely to be dissatisfied overall than those who did not.

Note that a similar kind of analysis wouldn't be so readily feasible using frequently used alternative approaches based on structural linear equation models without the development of two independent models.

6.7. Conclusions and further research

6.7.1. Conclusions

In this chapter we have presented the Bayesian network approach to traditional customer satisfaction research, in which we studied the relevance of three different service dimensions for overall customer satisfaction. Let us review the proposed approach in terms each sub-goal in more detail.

1. How can Bayesian networks be applied in service feature/dimension importance/performance study?

a. First, we demonstrated how Bayesian networks could be applied in service dimensions analysis for identifying the derived importance of service dimensions for overall (dis)satisfaction judgments,

At the beginning, we have dichotomised the overall satisfaction variable into three levels: low, medium, and high.

Since satisfaction at the dimension level was not operationalized by the customer questionnaire, variables reflecting service dimensions were created by inferring their values by k-means clustering algorithm based on satisfaction with specific features within the dimension. These variables were binary. We found that the clusters were well separated and could be considered as groups of high and low satisfaction. For these new variables, for each case in the dataset we assigned a value reflecting the level of satisfaction with the service dimension.

For both high and low level of overall satisfaction, we expressed their probabilities in terms of probability of high and low level with satisfaction with each dimensions, respectively. The procedure we proposed for this purpose is based on the one-way sensitivity analysis in the model, in which the dependent probability is the probability of high, and low, overall satisfaction, and the parameters are marginal probabilities of high and low satisfaction with each feature. We showed that this probability could be illustrated graphically with linear functions.

These graphs confirm the findings in [Mittal *et al.*, 1998] in that they show the diverse nature of the influence of satisfaction with a feature on overall service satisfaction: low levels of satisfaction are found hardly sensitive to dissatisfactory experiences with service dimensions, whereas high overall satisfaction shows in this respect an increased dependence.

We concluded that customer service could be classified as satisfier/dissatisfier. Similarly, we can classify billing also to the same; nevertheless, billing quality has a more substantial impact on satisfaction than customer service has. Tariffs, due to their positive impact on moderate satisfaction and negative impact on high satisfaction, warrant a closer look to arrive at the right conclusion.

b. Furthermore, we have developed a procedure and evaluated Bayesian networks with regard to supporting marketing decisions by means of importance/performance analysis.

Based on the strength of the influence we classified the service dimensions into categories of importance, and augmented with their performance, we carried out the analysis of priorities for improvement.

In order to calculate the performance of the service dimensions, we compared their marginal probability distributions with the one for overall satisfaction, and we found that the performance of all the three dimensions can be classified as low.

From the importance-performance analysis it follows that the company should undertake some actions to improve the performance of the considered service aspects billing being the first priority, and customer service being the second. Further insight regarding phone tariffs is required to formulate a relevant marketing policy in this respect.

We conclude that the approach can be used for importance/performance analysis concerning service dimensions.

c. Finally, we examined Bayesian networks in terms of discovering interaction effects (synergy and negation) among service dimensions.

We have found it likely that some potential determinants of overall satisfaction do not manifest an apparent influence when considered apart from other factors. It can however at the same time happen to be an important factor catalysing the impact of other service dimensions. Synergy effects that can be observed in this situation may be either positive or negative. Therefore, we included a study of interaction effects among the dimensions.

The procedure we proposed is based on the two-way sensitivity analysis in the model, in which the dependent probability is the probability of high, and low, overall satisfaction, and the parameters are marginal probabilities of high and low satisfaction with each feature.

For instance, we have observed a strong positive synergy between satisfaction with customer service and invoicing, and negative effect between invoicing and tariffs.

We can conclude also that the Bayesian network approach is very useful in determining interaction effects.

2. What are the strengths and weaknesses of Bayesian networks in terms of specific statistical and modelling issues, such as data distributional assumptions, missing data handling, etc.

Regarding the strengths we have found that they can allow for the use of all data in one analysis, and allow for deeper investigation of relationships.

We have been able to discover most importantly that Bayesian networks can easily account for all the data in one model in situations in which other methods would require the use of multiple models. For instance, in most often used covariance-based SEM modelling, it is not so easy to study the moderating effects [Gefen *et al.*, 2000]. The multi-sample approach recommended in such situations [Jöreskog and Sörbom, 1989] requires that the parameter estimates of the same model be examined by running the analysis on distinct sub-samples, and testing for the difference in χ^2 statistics obtained for the two models. We acknowledge that in the case of the least squares-based models, such as regression and PLS, the analysis of interaction effects is much more straightforward [Neter *et al.*, 1990].

Furthermore, in the classical approach to feature performance analysis, factor analysis is followed by regression analysis [Naumann and Giel, 1995; Oliver, 1996]. Factor analysis is used to construct and operationalise satisfaction at a higher, dimensional level of abstraction based on perception of the specific service/product features. Some features can be tested against their relevance and, possibly, excluded from the study as not "loading" on the dimension, thus non-relevant. Afterwards, linear relationships between each dimension and overall satisfaction are examined using regression analysis. In comparison to the above approach, the presented methodology enables deeper investigation of relevance of dimensions at various levels of the general performance. Furthermore, these relationships can be represented with informative charts for easier interpretation.

As far as the weaknesses are concerned, those weaknesses that we encountered in this case study have already been discussed in the previous case studies.

6.7.2. Managerial implications

We can recommend that managers can apply the methodology to classify service dimensions. In this study we concluded that customer service could be classified as satisfier/dissatisfier. Similarly, we can classify billing also to the same; nevertheless, billing quality has a more substantial impact on satisfaction than customer service has.

From a managerial perspective, outcomes of the present technique seem to be of interest, as they indicate which dimensions should be taken care of, and which of them are less important and deserve less attention. For instance we found, that the company should undertake some actions to improve the performance of the considered service aspects billing being the first priority, and customer service being the second. Further insight regarding phone tariffs is required to formulate a relevant marketing policy in this respect.

All the relationships are viewed probabilistically, thus allowing for easy interpretation. The outputs of this analysis are of a probabilistic nature and easy to interpret for managers.

It is also possible to find out the synergy and negation effects, if exist, between perception of different service dimensions.

Note that a similar kind of analysis wouldn't be so readily feasible using frequently used alternative approaches based on structural linear equation models without the development of two independent models.

6.7.3. Limitations

It is important to point several topics limiting the usability and generalizability of research presented in this chapter.

A well-known problem that occurs in traditional customer satisfaction studies is that if a list of features included in the investigation becomes too long, then it makes the analysis complicated and unreliable. The models require in this situation too many parameters that cannot be reliably estimated with available data. Alike, one of the limitations of the presented approach is that it is also not feasible to study the interaction of many dimensions at the same time. Since the conditional probability table is growing very quick with the number of features, and yielding nuisance with the model's parametric estimation.

Satisfaction with dimension was created artificially by finding two clusters of users in terms of their satisfaction with features relevant to each dimension. We should test how the presented technique will perform if satisfaction with dimensions is also operationalized by the questionnaire.

6.7.4. Further research

With respect to further research, a number of issues can be addressed to corroborate usability of the presented approach theoretically as well as for marketing practice. Predominantly, future research may be focused on investigation of models involving more dimensions and testing sensitivity of the approach in this respect.

7. Case study 4: Classification of features

7.1. Introduction

In the previous chapter we illustrated the use of the Bayesian network approach for gaining insight into the nature of the relation between the satisfaction at the service dimension level and the overall customer satisfaction. The abstract service dimensions are however difficult to control and manage in practice, because they usually encompass a wide range of specific and diverse service/ product attributes. The practical applicability of results of such studies is therefore limited. The predominant purpose of practical satisfaction research should be thus to evaluate the importance and performance of service/product attributes, rather than service/product dimensions, with relation to overall customer satisfaction. To avoid confusion, as a service/product attribute we will define "any aspect of the product itself or its use that can be used to compare product alternatives" [Grunet, 1989]. We will interchangeably with a word "attribute" use also the word "feature". This assessment of the importance and performance boils generally down to the classification of the nature of relation between each feature and the overall satisfaction score and can be defined in a way that we proposed in Chapter 6. Recall that we have then defined four kinds of features' nature: sastisfier/dissatifier, exciter, basic, and non-relevant. So, our focus in this chapter is to classify the features.

Consequently, research design in the feature performance studies requires that we assess direct relationships between satisfaction with service features and overall satisfaction. In practice there are however various well-known conceptual and practical difficulties involved in such a design. They both come from the fact that the number of features in such studies is usually too large. Quite often this number reaches hundred features or more. This means that we have to deal with one dependent variable, i.e. overall satisfaction, and a lot of independent variables.

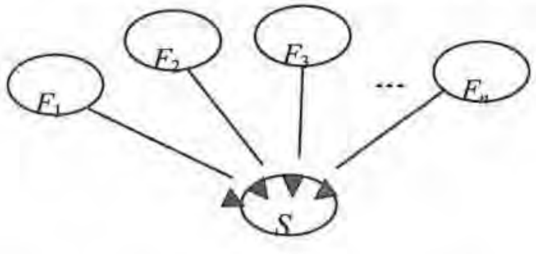


Figure 7.1.1 Initial model for modelling the joint probability distribution in satisfaction studies.

The initial model for modelling the joint probability distribution for overall satisfaction and service features is shown in Figure 7.1.1. This model assumes that the features (F_i) are independent form each other, but they become dependent when a value of overall satisfaction (S) is known. The approach in which all service features are parents of overall satisfaction should be, nonetheless, abandoned for two reasons.

Firstly, such an approach should be neglected on the basis of its conceptual shortcomings. It is basically rather unacceptable to presume that the ultimate overall satisfaction judgment can be a result of simultaneous processing of all the features identified in the study. It is hardly imaginable that satisfaction is a result of processing information from so many equally important features. Indeed, findings in psychological disciplines report that the span of absolute judgment and the span of immediate memory impose severe limitations on the amount of information that humans are able to receive, process, and remember. In one of the most classical articles in the history of psychology, George A. Miller [1956] showed experimentally that this span of attention oscillates around the "magical" number of seven. More interestingly, he argued that humans manage to break this informational bottleneck by organizing the stimulus input simultaneously into several dimensions, and subsequently into sequences of chunks [Miller, 1956].

Secondly, such an excessive number of potential service features is usually much bigger than can be taken into account in the analysis due to computational problems, since it would require a very large database of cases to estimate the parameters probably of any available statistical technique to date. This problem is quite common also in faithful estimation conditional distributions in case of Bayesian network models, as is the case for each possible configuration of states of service features (causes) as parents of overall satisfaction (effect). A too small dataset can affect the reliability of the parameters, and bias the end results. The amount of attributes in a study affects also the complexity of interpretation. This is in fact a common problem in practical satisfaction studies [e.g., Oliver, 1996].

In the Bayesian network modelling literature, several authors [e.g., Henrion, 1987; Olesen *et al.*, 1989] have proposed various methods to ease that problem including 1) parent divorcing [Olesen *et al.*, 1989], and 2) noisy functional dependencies [Good, 1961; Pearl, 1989]. We speculate that application of these two techniques can be valuable also in practical customer satisfaction studies.

We address the technique of parent divorcing more formally in Section 7.3. As a way of introduction, it is now worth to mention that it implies that new random variables are introduced as effects of service features and parents of overall satisfaction. In this way, we obtain a new model of overall satisfaction, that we will call a *mediated* model. These new variables in our model can presumably reflect customer satisfaction with a relevant service/product dimension. This approach suffers however from imperfections. The most important of them is that it would require measuring customer satisfaction with respective dimensions by the survey. This results in the extension of the question list by another six-ten questions. There are known important aspects that affect the quality of satisfaction research when too many questions are involved in a study. For instance, Douglas [1995] argues that when a questionnaire is too long respondents get tired of answering questions and are not willing to participate further, which is known as the phenomenon of response fatigue. This contributes to dropout. Furthermore, they tend to give uninvolved answers that are not a true reflection of their actual respondent's standpoint, which is another threat to the quality of the research. Last but not least, asking each additional question on a questionnaire is usually an extra cost for a company that orders the customer satisfaction study.

As regards the noisy functional dependencies, typical examples in the literature include noisy-OR, noisy-AND, noisy-MAX, and noisy-MIN models [Good, 1961; Henrion, 1987; Srinivas, 1993]. They belong to the class of models known as models of independence of causal influence (ICI), or models of causal independence [Heckerman, 1993; Heckerman and Breese, 1994]. An advantage of these models is that they require less parameters to be estimated, at the expense, however, of simplifying the conditional probability $P(S \mid F_1, F_2, \ldots, F_n)$. Of these methods, our initial examination concerns the noisy-OR model since it seems to be the most widely applied model in this family. We treat the noisy OR-gate in more detail in Section 7.5.

7.1.1. Objectives

This chapter is the second one in this dissertation aimed at the problem that we sketched in the research question no. 4 in Chapter 1. Recall that we have then formulated this problem as "How can BNs be applied in service feature/dimension importance/performance study?" In the previous chapter, we analysed a methodology to assess importance of service dimensions. In the current chapter, the main objective is to adapt and examine Bayesian networks in classification of service features (attributes).

In particular, this study has clearly three sub-goals:

a. to evaluate the mediated model of overall satisfaction based on the technique of parent divorcing in the analysis of feature importance,
b. to find out whether in the mediated model, it is possible to treat satisfaction with service dimension as a hidden node, and thus optimise a questionnaire by not asking about satisfaction with service dimension,
c. to evaluate the noisy-OR model of overall satisfaction in the analysis of feature importance.

Firstly, we evaluate the mediated model of overall satisfaction based on the technique of parent divorcing. More precisely, our primary mediated model contains customer satisfaction judgments operationalized by customer questionnaire, so that all variables are treated in this model as observed. Evaluation of this model will be achieved by two criteria. To begin with, we will assess the ability of performing classification of service features by counting the number of meaningfully classified features. Furthermore, we will validate the model for its ability of predicting overall satisfaction.

As an alternative to the primary mediated model with service dimensions, whose values have been directly operationalized, we consider also a model, in which all service dimensions are treated as hidden nodes. In these hidden construct model, all the necessary probabilities can be estimated on the basis of the remaining variables and the dependencies implied by the model by means of an optimisation technique, such as the EM algorithm. Therefore, secondly, another important question addressed in this study is whether it is necessary to measure satisfaction with a service/product dimension by asking it directly in a customer questionnaire, or, alternatively, whether it is possible to optimise a questionnaire by not asking about satisfaction with service dimension, and by deriving it from the scores of satisfaction with features and overall satisfaction. In order to judge it, we compare alternative models of customer satisfaction: 1) one in which the level of satisfaction with a service dimension is inferred indirectly from data by maximum likelihood estimation, and 2) another one in which this level is explicitly measured by the data questionnaire and taken account of. To find out whether the two models are equivalent in practical satisfaction study, we will perform two types of comparative validation: 1) in qualitative validation, we will compare the results of classification of the features using the scale developed in the previous study, 2) in addition, the predictive accuracy will serve as the second type of validation.

Thirdly, this chapter is aimed at evaluating other distributional dependencies, more specifically the noisy OR-gate, for the analysis of feature importance. We will assess the ability of performing classification of service features by counting the number of meaningfully classified features.

7.1.2. Organisation

The remainder of this chapter is organized as follows. First, in Section 7.2, we describe in more detail the questionnaire and data collected in a customer satisfaction study that serves as a background for our discussion. Then, we introduce formally the technique of parent divorcing in Section 7.3, cast light on feature selection in Section 7.3.1, discuss model specification and estimation in Section 7.3.2, and perform predictive validation in 7.3.3. Section 7.4 contains the results of feature classification with parent divorcing. Section 7.5 is entirely devoted to classification of features with the noisy OR-gate. We close with conclusions in Section 7.6.

7.2. Data

Compared to the case study in Chapter 6, the current study requires a dataset, in which not only the data on overall satisfaction, but also directly on service dimensions are present. The data that we use in this chapter come from a study of customer satisfaction and commitment among companies that are customers of a bank company offering an infrastructure and end-user terminals for electronic payments in Belgium. The survey was prepared and administered by a Belgian market research agency in spring 2002. The respondents that took part and responded positively to the invitation included personnel of small outlets, such as shops, cafes, restaurants, and filling stations. The aim of the survey was to assess satisfaction with various aspects of the services and products used for electronic transaction systems. The collection of data resulted in a sample of 1201 respondents.

7.2.1. Questionnaire

The questionnaires contained questions concerning overall satisfaction, perceptions of products and other service attributes linked with the use of the terminals. Originally, the questionnaire contained a list of nine dimensions, including: image, price, quality, products, sales service, maintenance, billing, administrative services, and communication and promotion. In each section devoted to these dimensions, the questionnaire included a listing with varied quantities of specific service features. This number varied form six to ten features per dimension. In total, about 80 distinct features were represented with the questionnaire.

Construct	Symbol	Items
Overall Satisfaction	Sat	Can you tell me to what extent are you overall satisfied with products and services of ...?
Image *Features:*	Im	Generally considered, how do you find the image of ...?
1. Trust	Im1	... is a company that I trust.
2. Customer-orientation	Im2	... is a customer-oriented company.
3. Technology	Im3	... is a technologically advanced company.
4. Security	Im4	... is a company that cannot be surpassed concerning the security of the products.
5. Competency	Im5	... is a company with a competent personnel.
6. Competitiveness	Im6	... is a company that has a competitive position in the market.
7. Grow	Im7	A company that will certainly grow in the coming years.
8. Professionalism	Im8	A professional company.
Product Quality *Features:*	PQ	In general, how do you evaluate the quality of the products?
1. Top technology	PQ1	The products are technologically on the top.
2. Durability	PQ2	The products are durable.
3. User-friendliness	PQ3	The products are user-friendly.

4. Information	PQ5	The products supply enough information.
5. Speed	PQ7	The products go fast.
6. Accessibility	PQ9	The accessibility to the network is big enough.
Billing *Features:*	B	How do you in general perceive the invoicing of ...?
1. Clarity	B1	The invoices are clear.
2. Frequency	B2	The invoices should be made more often than four times a year.
3. Conditions	B3	The invoicing respects the conditions fixed in my contract.
4. Correctness	B4	The invoicing is always 100% correct.
5. Amount	B5	There are too many different invoices with respect to the products I use.
6. Design	B6	The design is pleasant.
Price *Features:*	Pr	How do you in general perceive the price policy of ...?
1. Hardware prices	Pr1	The prices that ... employs are honest.
2. Transaction prices	Pr2	The price level of transactions is fair.
3. Maintenance prices	Pr3	The price that you pay for maintenance is fair.
4. Rent prices	Pr4	The price level of the rent of terminals is fair.
5. Transparency prices	Pr5	The prices are easily comprehendible and transparent.
6. Customisation	Pr6	The prices are tailored to the profile of the user.
Communication & Promotion *Features:*	CP	In general, how do you perceive the communication policy of ...?
1. New products	CP1	... informs its clients in a good way about its new products.
2. Brochure	CP2	The brochure ... contains useful information.
3. Advertising	CP3	The way in which ... advertises in radio and posters is good.
4. Availability	CP4	It is easy to get promotional materials for the shop.
5. Amount	CP5	There are enough promotional materials for the shop.
6. Attractiveness	CP6	The promotional material for the shop is attractive.
7. Mailing	CP7	I get too many mails because of ....
8. Language	CP8	... informs its clients in their language.

Table 7.2.1 Operationalisation of constructs (incomplete list).

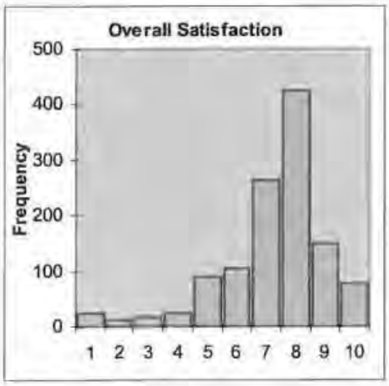
Overall customer satisfaction was measured only with one question that should be answered with a number from 1 to 10, where 1 stands for "extremely dissatisfied," and 10 for "extremely satisfied." Satisfaction with service dimensions was measured on two levels: 1) on the level of overall satisfaction with that service dimension, 2) on the level of specific features that measure a precise aspect of the dimension. On the former level the constructs were measured with one item with the responses in the range of five values ranging from "very bad" to "very good". Additionally, in each section dedicated to each service dimension a list containing specific attributes was asked, the perception of which was measured with the 5-point Likert scale, ranging from "strongly disagree" to "strongly agree." Respondents were also given the possibility to select the answer "do not know." The dimension Quality did not actually measure satisfaction, but was meant to find out whether a customer experienced problems with usage, and whether these problems were solved. This was a reason not to include this section into further analysis.

Last but not least, we must make it clear that the one-item measurement instruments that the questionnaire contains cannot be validated against reliability. We therefore need to make an assumption that the measurement error is minimal, and the observed responses reflect true scores.

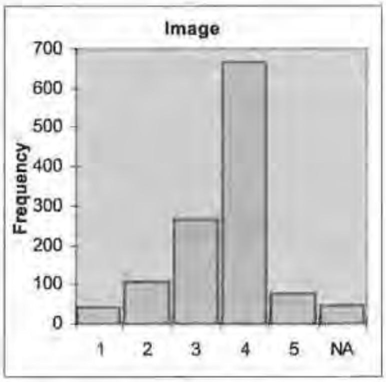
7.2.2. Data preparation

In Figure 7.2.1 we present histograms with frequencies of the responses on all of the nine dimensions present in the questionnaire. We do not present the histograms for features since there are too many of them. For sales service as a dimension, 495 responses were not available, and more than 1000 on all sales service features. The same applies to administrative services (286 cases miss values on the dimension, 921-1099 cases have no data on features). Similarly, for all features of maintenance, more than 50% of responses were missing. For these reasons, we have decided not to take these service dimensions into account.

a)

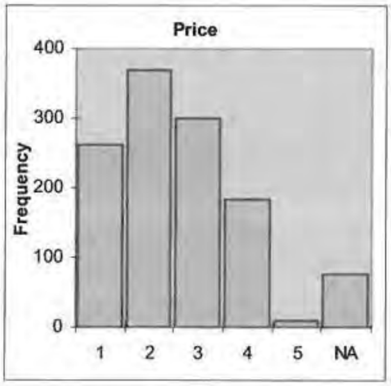


b)



c)

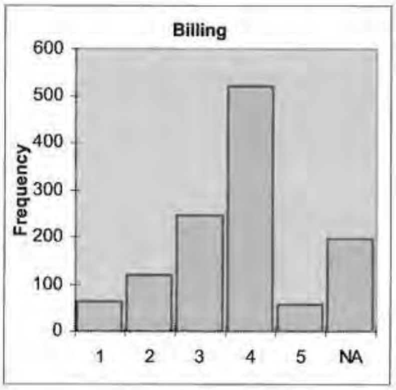
d)

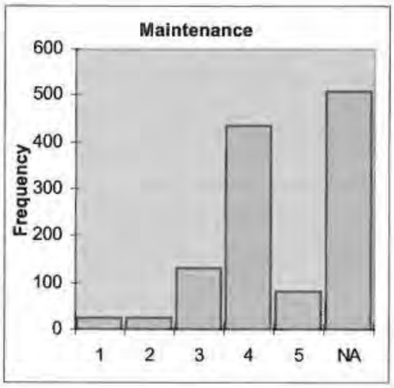

Price
Frequency
400
300
200
100
0
1
2
3
4
5
NA


Product quality
Frequency
900
800
700
600
500
400
300
200
100
0
1
2
3
4
5
NA

e)

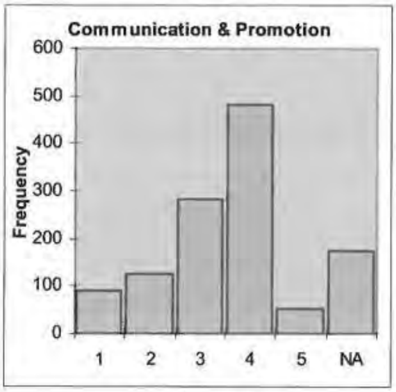
f)


Billing
Frequency
600
500
400
300
200
100
0
1
2
3
4
5
NA


Maintenance
Frequency
600
500
400
300
200
100
0
1
2
3
4
5
NA

g)

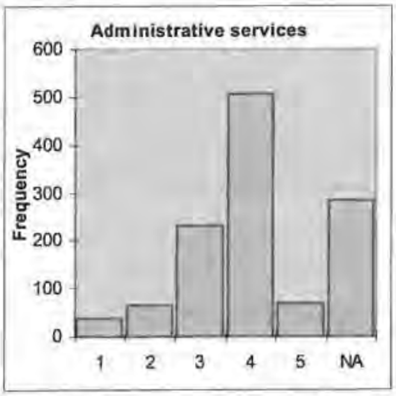
h)


Communication & Promotion
Frequency
600
500
400
300
200
100
0
1
2
3
4
5
NA


Sales service
Frequency
600
500
400
300
200
100
0
1
2
3
4
5
NA

i)

j)



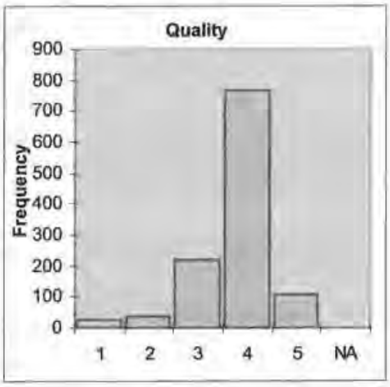


Figure 7.2.1 Histograms of a) overall satisfaction, b) image, c) price, d) product quality, e) billing, f) maintenance, g) communication and promotion, h) sales service, i) administrative services, j) quality. (Note: "NA" represents a missing datum.)

By comparing the histograms for features and for their relevant dimensions, we could notice that there is generally higher proportion of people who are highly satisfied with specific service features than the proportion of people who are satisfied with the service dimension. For instance, there are 74 respondents who are very satisfied with image (they have responded with "very good"), whereas the average number of respondents who are very satisfied with specific image features is 195. The respondents tend thus to express more extreme satisfaction with features than with dimensions. Apparently, respondents tend to evaluate the specific features usually higher than the dimension itself in general. Respondents are less radical and more careful in making extreme judgments about more abstract concepts like dimensions than features.

Subsequently, we have collapsed the number of possible values that each concept has been operationalized with, so that each variable could take one of three values, instead of ten for overall satisfaction, or five, as was the case for service dimensions and features. For this purpose, we performed the grouping of values by assigning a new value to the values reported in the data set. With regard to the features we grouped together the three lowest response values starting from "completely disagree", while the responses on service dimensions expressed as "vary bad", "bad", and "neither bad or good" received a common value that we called "bad". Overall satisfaction was recoded accordingly, i.e., values greater than 7 received the value "very good", from 5 to 7 "good", and the other values were relabelled to "bad". Furthermore, we have recoded the responses valued "don't know", and other empty entries into missing values. We allowed for the missing values, and we have hold all the 1201 cases for the analysis.

7.3. Introducing mediating variables with divorcing

In order to handle the parameterisation of models with so many parents, as is the case in feature performance analysis, we can pool some of the features together and introduce mediating variables. This approach is in the Bayesian network literature known as *parent divorcing* [Andreassen *et al.*, 1989; Jensen, 2001]. Formally, if we have a set of nodes F_1, F_2, ..., F_n which are parents of node S (see Fig. 7.1.1), we can divorce the set of parents F_1, ..., F_i from the parents F_{i+1}, ..., F_n by introducing a mediating variable D, making D the child of the nodes F_1, ..., F_i and a parent of S. This approach can be shown graphically in Figure 7.3.1.

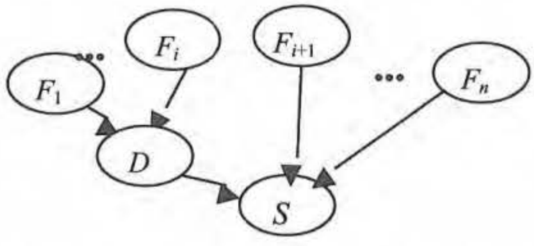


Figure 7.3.1 The nodes F_1, ..., F_i are divorced from F_{i+1}, ..., F_n by introducing the mediating variable D.

Of course, the question that now arises is whether the joint probability distribution $P(F_1, F_2, \ldots, F_n, S)$ encoded by the two models, i.e., before and after divorcing, stays the same and are equal. Our objective is naturally that it stays the same as before. The assumption on which the two joint distributions are the same is based on the following reasoning. Let us denote by r_i the number of states of the variable F_i, and by f_{ik}, $k=1,\ldots,r_i$, a particular instantiation of the variable F_i. Let us assume that the divorcing variable D has m states d_1, ..., d_m. Now, the idea is to lump different configurations of parent variables $(f_{11}, \ldots, f_{i1})$ and $(f_{11}, \ldots, f_{i2})$ together in one state d_i if and only if $P(S \mid f_{11}, \ldots, f_{i1}, F_{i+1}, \ldots, F_n) = P(S \mid f_{11}, \ldots, f_{i2}, F_{i+1}, \ldots, F_n)$. In this way, we can partition the set of configurations F_1, ..., F_i into the sets d_1, ..., d_m. It must be noted that when these conditions are met, one does not need additional data on the mediating variables to quantify the conditional dependencies for the mediating variable D given the parent nodes F_1, ..., F_i, and for the effect variable S given D, F_{i+1}, ..., F_n – these dependencies can be determined based on the distribution $P(S \mid F_1, F_2, \ldots, F_n)$. Similarly, we can divorce more parent variables by introducing more mediating variables, and making them parents of the variable S.

Based on these aforementioned assumptions divorcing works best. How can it be applied in our problem? Let us assume that the parent variables F_1, F_2, ..., F_n are service/product features, S is overall satisfaction, and D_i are some mediating variables.

A natural selection for the mediating variables D_i are customer judgements concerning satisfaction at a service dimension level. However, there still remain two problems to be tackled. Firstly, the original number of features per dimension will be very difficult to retain. For example, if we retain six features, e.g., F_1, F_2, ..., F_6, each of them ternary, that are supposed to concern one specific dimension, then there are 3^6=729 different configurations of their specific values. That is much too much, given the size of our dataset (1201 cases), to estimate the conditional probabilities $p(D_i| F_1, F_2, \ldots, F_6)$. If we assume that the features are binary variables, then we still have 2^6=64 combinations. Similarly, if we keep hold of three features per dimension, then they make up a set of 27 different combinations of values. Regardless how many states r_{D_i} the mediating variable should be defined with, it could still be difficult to estimate all the probabilities required to parameterise the dependency between service dimension and features.

Secondly, it is rather unrealistic to assume that we can decrease the number of configurations of service features, e.g., from 729, or 64, down to a number of states r_{D_i} of the mediating variables of a satisfying level. We must take two issues into account when deciding upon the satisfying number of states for the mediating variable. First of all, it may not be too big since the number of states of the mediating variable that makes it efficient to estimate the conditional probability table of overall satisfaction given the mediating variables. Next, it may not be too small, since then it is unrealistic to assume that many configurations will fulfil the condition $P(S\,|\,f_{11}, \ldots, f_{i1}, F_{i+1}, \ldots, F_n) = P(S\,|\,f_{11}, \ldots, f_{i2}, F_{i+1}, \ldots, F_n)$. Lastly, we would require that the service features measure perfectly all the aspects of the dimension, which is supposed to be measured with the question about overall satisfaction with a respective service dimension.

It follows that the approach with introducing satisfaction at the dimension level as mediating variables, as pictured schematically in Fig. 7.3.1, must be used with caution. However, since we are in the first place interested in the ultimate classification of features, and in the second place with a faithful modelling of the joint probability distribution $P(F_1, F_2, \ldots, F_n, S)$ then we assume that the variables in the intermediate layer can be approximated by customer judgments on satisfaction with the service dimensions, and then this approach may be acceptable.

The model that we evaluate is our proposal for reduction in the complexity of the first model. It contains three intermediate hidden nodes. These nodes represent customer satisfaction with a certain service dimension as a whole. In this model, the overall satisfaction is indirectly dependent on service features through satisfaction at a dimension level.

Finally, we should check if the assumptions are fulfilled. It is likely that the joint probability distribution is now different from the distribution before. In general, one way is to evaluate the distance between the two distributions $P(S\,|\,f_{11}, \ldots, f_{i1}, F_{i+1}, \ldots, F_n)$ and $P(S\,|\,f_{11}, \ldots, f_{i2}, F_{i+1}, \ldots, F_n)$. It is however not possible since we do not have the original joint distribution $P_{oryg}(F_1, F_2, \ldots, F_n, S)$, because it is too big and does not fit in the memory.

7.3.1. Feature selection

Because the original number of features per dimension was too large for reliable estimation, we have decided to select only three service dimensions with three features per dimension for further analysis. From the five dimensions remaining after data inspection described above, we selected randomly three service dimensions: Price, Product Quality and Image. This decision has the consequence that, on average, each conditional distribution $p(D_i| F_1, F_2, \ldots, F_6)$ is estimated on basis of $1201/27 \approx 44$ data points. In reality, it can be more, or less, depending how many cases there are for a specific configuration of features' values.

Next, we have done the same random selection of features per dimension. For Price policy, we have chosen features that refer to 1) the prices of the terminals and other hardware (Pr1 – Hardware Prices), 2) prices of the transactions carried out by the terminals (Pr2 – Transaction Prices), and 3) the easiness to comprehend and the transparency of the price scheme (Pr5 – Transparency Proces). Specific features that relate to the Image of the company are: 1) trust towards the company (Im1 - Trust), 2) perception of its technological advancement (Im3 - Technology), 3) the security of the products offered by the company (Im4 - Security). The last selected group of features are meant to capture the specific aspects of the Quality of products, and especially terminals. We have chosen the following features in this respect: 1) the products present top technology (PQ1 – Top technology), 2) the user-friendliness of the products (PQ3 – User-friendliness), and 3) the products supply enough information on screen (PQ5 - Information). The precise operationalization of each concept used in the study is shown in Table 7.2.1.

We shall assume that overall satisfaction with service is directly determined by customer experience with the service. The experience with the service concerns customer perceptions of the service dimensions, which are in turn measured by perceived performance of specific attributes, along which the customers view the service. The network structure is fixed *a priori* and corresponds to the belief that (dis)satisfying service dimensions cause overall satisfaction. Furthermore, we treat the service features as measures of how (dis)satisfying the corresponding service dimension is.

7.3.2. Model specification

Figure 7.3.2 presents the graphical representation of the models used in the study. We consider two models. In one of them, nodes representing dimensions are treated as hidden; we call this model "est". In the other one the service dimensions are treated as other nodes for which there are observations; this model is referred to as "obs".

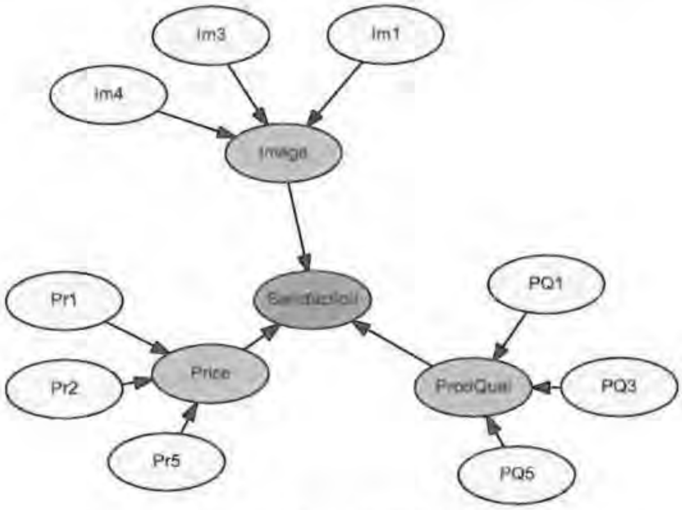


Figure 7.3.2 Structural model evaluated in the study.

The intermediate nodes in the model with hidden dimensions represent latent constructs, i.e. the constructs that refer to no data, and hence are treated as hidden nodes in the Bayesian network terminology. As a result, the parameterisation of the conditional probability tables in this model must be performed by means of an approximation technique. As it was the case in the study in Section 5.2, the values of the conditional probabilities were supposed to maximize the loglikelihood of the model. In other words, we have strived to find the maximum likelihood (ML) estimates of the probabilities. For this purpose we have applied the EM algorithm, described in more detail in Section 5.3.4, with the parameter priors set to zero in order to decide the data alone about the values of the resulting probabilities. From 100 runs of the algorithm, each with different random initialisations of parameters, we have chosen one that yielded the highest value of the loglikelihood. We have set the value of the convergence threshold between the two consecutive iterations of the algorithm to 0.0001, which should be a sufficiently small value in this respect.

As a consequence of the Maximum Likelihood optimisation, we have faced the problem of aliasing of the states for service dimensions [e.g., Chickering and Heckerman, 1997], since maximum likelihood estimates are not sensitive to the actual labels of values that hidden variables take. Although it is not necessary to tackle this problem for the purpose of investigating the effects of features on overall satisfaction, it can give some indications on the quality of the estimated model. We have found that the hypothetical true meaning of the states has been difficult to establish. As the rule of thumb, we have strived to label the unknown states by performing inference in both models, so that the marginal distributions for each service dimension in both models reflect the possibly most analogous pattern.

7.3.3. Validation

Both models have the same number of structural parameters. The size of the conditional probability table of overall satisfaction is $3^4 = 81$, as well as of service dimensions, which means that we should estimate 27 distinct conditional probability distributions.[1] For each combination of service features instantiations we have thus on average about 44 cases from which we can estimate three probabilities, among which only two are non-redundant.

The predictive validity has been applied to measure the fit of the model to the data. We have treated the overall satisfaction as the target node, whose value should be determined on the basis of the values of all twelve other variables. We have performed 5-fold cross validation of each classifier. Table 7.3.1 below shows the averages for various scores of predictive validity for both models.

	Model *est*	**Model *obs***
Cases verified	1188	1188
Cases classified correctly	771	754
Classification accuracy	64.9%	63.46 %
Log loss	0.7240	0.7787
Brier score	0.1964	0.2089

Table 7.3.1. Classification accuracy for both models in the study.

The results show slight difference in performance between both models in favour of model with dimensions estimated from the data. The classification accuracy of 64.9% for this model between the two models in terms of their classification power, and consequently both models could be interchangeably used in classification tasks. The accuracy of around 65% should not be seen as high however but the practice shows that in satisfaction studies higher accuracy is rarely reached.

When we take account of the uncertainty of the classification decision coming from the posterior probability distribution over the target node, expressed with the Brier score, we see that the Model est with its score of 0.1964 performs somewhat better then Model *obs*, whose score amounts to 0.2089. We could therefore say that Model *est* is closer to an optimal classifier than Model *obs*.

[1] We assume universally in this work that the parameters that belong to different instantiations of parents are independent (local parameter independence.) See Section 4.2 for more details.

7.3.4. Marginal probabilities

We will begin with the marginal probabilities for both models to get an initial idea about the differences in parametric estimation between the two models.

		Model *est*	Model *obs*
Satisfaction	low	0.060	0.060
	med	0.393	0.395
	high	0.547	0.544

Table 7.3.2 Marginal probabilities for Overall satisfaction yielded by both models.

From Table 7.3.2, we can see that there are practically no differences in the values of marginal probabilities for Overall Satisfaction, which is a positive, but not surprising result.

	Image	Price	Product Quality
low	0.369	0.864	0.202
med	0.569	0.128	0.679
high	0.060	0.007	0.119

Table 7.3.3 Marginal probabilities for dimenions yielded by model obs.

	Image	Price	Product Quality
low	0.279	0.486	0.215
med	0.369	0.258	0.397
high	0.352	0.256	0.388

Table 7.3.4 Marginal probabilities for dimenions yielded by model est.

However, when we examine the marginals for the dimensions, we see that significant differences exist between values in Table 7.3.3 and 7.3.4. Low satisfaction with Price amounts to 0.864 for model *obs*, whereas for model *est* it is only 0.486. Generally, every other pair of corresponding probabilities assures us that the mediating variables in model *est* represent, in fact, some other concepts, distinct from the variables representing satisfaction with dimensions operationalized by the questionnaire and present in model *obs*. This finding can also be a signal that conditional probability tables of the two models are dissimilar. We have found that service features have virtually the same marginal distributions, so we do not find the need to include these distributions in this discussion.

7.3.5. Conditional probabilities

This initial examination of the tables is valuable for gaining idea of the local dependencies between features and service dimensions. We do not present the complete conditional probability tables for the dimensions and overall satisfaction since the size of these tables, amounting to 27 independent distributions, is too big. Apart from the presentation problem, the tables would be quite difficult to analyse due to the information overflow. Instead, we show marginal distributions of dimensions given evidence only on one feature at a time. The tables below are obtained by performing inference in the Bayesian network by instantiating the evidence nodes, i.e., the service feature nodes, updating the beliefs, and reading off the marginal probabilities for service dimension nodes.

Image

Let us take a look at Tables 7.3.5-7 with the dependencies for Image given Image features in the model in which all variables shown in Fig. 7.3.2 are treated as observed.

	Im1		
Image	low	med	high
low	0.610	0.300	0.214
med	0.381	0.653	0.626
high	0.008	0.047	0.160

Table 7.3.5 Marginal conditional probabilities for Image given Im1 for model obs.

	Im3		
Image	low	med	high
low	0.455	0.336	0.331
med	0.495	0.608	0.581
high	0.050	0.055	0.088

Table 7.3.6 Marginal conditional probabilities for Image given Im3 for model obs.

	Im4		
Image	low	med	high
low	0.364	0.337	0.475
med	0.583	0.596	0.462
high	0.053	0.067	0.063

Table 7.3.7 Marginal conditional probabilities for Image given Im4 for model obs.

We can notice that marginal distributions of Image are remarkably similar for medium levels of satisfaction with each of the three features. When respondents are dissatisfied with image features, the distributions are different. For instance, low satisfaction with Image is probable with 0.61 given dissatisfaction with Im1, whereas low satisfaction with Im3 yields the same level of satisfaction with Image with probability of 0.455. From the three features, the least chance that satisfaction with image is low can be found for Im4, as only 36% of respondents who are dissatisfied with Im4, are at the same time dissatisfied with Image.

With regard to the impact of features on high satisfaction with dimensions, we can observe that this impact is very low. Surprisingly, the distributions for moderate and high Image satisfaction for each feature (especially for Im1 and Im3) are not so different, which could indicate that feelings of satisfaction with the dimension remain at the same level regardless whether satisfaction with features is moderate or high.

Moreover, we can observe a clearly positive association between Im1 and Im3, and Image. This means that the higher satisfaction with those features, the lower chance of low satisfaction with Image, and the higher chance of high satisfaction with Image. However, relationship between Im4 and Image seems to be complex. Notably, when satisfaction with Im4 is low, low satisfaction with

Image is probable as of 0.364; given high satisfaction with Im4, low satisfaction with Image grows up to 0.475.

Now, let us look at what are the conditional distributions for Price given satisfaction with one feature at a time for model *est* contained in Tables 7.3.8-10. The purpose of this investigation is to compare and assess the difference on conditional probabilities between the two approaches.

	Im1		
Image	low	med	high
low	0.645	0.167	0.060
med	0.193	0.450	0.412
high	0.162	0.382	0.527

Table 7.3.8 Marginal conditional probabilities for Image given Im1 for model est.

	Im3		
Image	low	med	high
low	0.379	0.234	0.250
med	0.165	0.413	0.131
high	0.456	0.353	0.618

Table 7.3.9 Marginal conditional probabilities for Image given Im3 for model est.

	Im4		
Image	low	med	high
low	0.245	0.287	0.345
med	0.450	0.352	0.205
high	0.306	0.361	0.449

Table 7.3.10 Marginal conditional probabilities for Image given Im4 for model est.

The most apparent difference that we can observe is much higher probability of higher level of satisfaction with Image in case of model *est* than it is for model *obs*. In Table 7.3.5, the probabilities of high perception of Image is 0.008, 0.047, and 0.160, given low, moderate, and high state of satisfaction with the feature Im1 (Trust), respectively. The equivalent probabilities for the model *est*, as can be seen in Table 7.3.8 are 0.162, 0.382, and 0.527, so we can see that the respective marginal conditionals that result from the EM estimation are significantly different from the ones that follow from the model with observed dimensions.

Product Quality

Relationships between features of Product Quality and this dimension are more in line with our expectations, although remarkable is that low satisfaction with Product Quality is lowest when satisfaction with features is moderate, and not low (see e.g., probability of 0.155 given PQ1, and 0.164 given PQ5). This is an unexpected result.

	PQ1		
ProductQuality	low	med	high
Low	0.256	0.155	0.211
med	0.682	0.725	0.523
high	0.062	0.119	0.266

Table 7.3.11 Marginal conditional probabilities for ProductQuality given PQ1 for model obs.

	PQ3		
ProductQuality	low	med	high
low	0.366	0.197	0.137
med	0.542	0.706	0.688
high	0.092	0.097	0.174

Table 7.3.12 Marginal conditional probabilities for ProductQuality given PQ3 for model obs.

	PQ5		
ProductQuality	low	med	high
low	0.257	0.164	0.240
med	0.643	0.735	0.579
high	0.099	0.101	0.181

Table 7.3.13 Marginal conditional probabilities for ProductQuality given PQ5 for model obs.

Let us compare these conditionals wth the conditionals obtained for ProductQuality for the model *est*, which can be found in Tables 7.3.14-16.

	PQ1		
ProductQuality	low	med	high
low	0.323	0.162	0.105
med	0.351	0.491	0.219
high	0.325	0.347	0.675

Table 7.3.14 Marginal conditional probabilities for ProductQuality given PQ1 for model est.

	PQ3		
ProductQuality	low	med	high
low	0.342	0.197	0.215
med	0.433	0.353	0.397
high	0.224	0.449	0.388

Table 7.3.15 Marginal conditional probabilities for ProductQuality given PQ3 for model est.

	PQ5		
ProductQuality	low	med	high
low	0.397	0.147	0.199
med	0.134	0.598	0.168
high	0.468	0.254	0.633

Table 7.3.16 Marginal conditional probabilities for ProductQuality given PQ5 for model est.

Based on the comparison between Tables 7.3.11-13 and Tables 7.3.14-16, it can be seen again that the conditional marginals for Product Quality in model *est* are too high for high satisfaction. In general, this observation can be a result of the tendency of the EM estimation to smooth the distributions of the latent construct given its parents. More precisely, the maximum likelihood estimation method tends to distribute the uncertainty of the probabilistic parameters across all the states of the latent construct; as a result, the estimated distributions tend to become closer to uniform distribution. More importantly, the character of the dependency is not preserved, i.e., some cells in model *obs* contain the least probability, while the same cell in model *est* show the highest probability in the distribution (cf., 0.097 in Table 7.3.12 and 0.449 in Table 7.3.15).

Price

As regards Price, we should note especially negative association between satisfaction with Pr1 and Pr2, and high satisfaction with Price. In particular, high satisfaction with Price as a dimension is probable as of 0.005 given low satisfaction with Pr1, and as of 0.001 given high satisfaction with that feature. Exactly the same phenomenon can be observed for Pr2.

	Pr1		
Price	low	med	high
low	0.914	0.678	0.786
med	0.081	0.307	0.213
high	0.005	0.015	0.001

Table 7.3.17 Marginal conditional probabilities for Price given Pr1 for model obs.

	Pr2		
Price	low	med	high
low	0.893	0.776	0.699
med	0.102	0.209	0.300
high	0.005	0.015	0.001

Table 7.3.18 Marginal conditional probabilities for Price given Pr2 for model obs.

	Pr5		
Price	low	med	high
low	0.895	0.849	0.777
med	0.097	0.150	0.200
high	0.008	0.001	0.023

Table 7.3.19 Marginal conditional probabilities for Price given Pr5 for model obs.

Finally, let us see what are the conditional distributions for Price given satisfaction with one feature at a time for model est contained in Tables 7.3.20-22. Similarly, from the comparison between the model *obs* and model *est*, we can make the same remarks concerning the significant difference in the conditional probabilities and the tendency of the EM-ML estimation to smooth the probabilities.

	Pr1		
Price	low	med	high
low	0.592	0.100	0.271
med	0.188	0.511	0.382
high	0.218	0.389	0.347

Table 7.3.20 Marginal conditional probabilities for Price given Pr1 for model est.

	Pr2		
Price	low	med	high
low	0.573	0.163	0.231
med	0.202	0.541	0.033
high	0.225	0.295	0.735

Table 7.3.21 Marginal conditional probabilities for Price given Pr2 for model est.

	Pr5		
Price	low	med	high
low	0.493	0.405	0.724
med	0.239	0.331	0.095
high	0.267	0.264	0.179

Table 7.3.22 Marginal conditional probabilities for Price given Pr5 for model est.

Overall Satisfaction

Finally, it is informative to consult the marginal distribution for Overall Satisfaction conditionally on different instantiations of service dimensions. These distributions, shown in Tables 7.3.23-25, suggest strong positive effect of satisfaction at dimension level on overall satisfaction. The only exception can be noticed when low overall satisfaction is more probable when satisfaction with Image is high (0.057), than as if it was moderate (0.025).

	Image		
Overall Satisfaction	low	med	high
low	0.116	0.025	0.057
med	0.456	0.367	0.224
high	0.419	0.607	0.718

Table 7.3.23 Marginal conditional probabilities for Overall Satisfaction given Image for model obs.

	Product Quality		
Overall Satisfaction	low	med	high
low	0.122	0.051	0.008
med	0.527	0.379	0.259
high	0.350	0.569	0.732

Table 7.3.24 Marginal conditional probabilities for Overall Satisfaction given Pr2 for model obs.

	Price		
Overall Satisfaction	low	med	high
low	0.070	0.001	0.001
med	0.431	0.166	0.124
high	0.499	0.833	0.875

Table 7.3.25 Marginal conditional probabilities for Overall Satisfaction given Pr5 for model obs.

In Tables 7.3.26-28, we report the values of conditional probabilities obtained with the model *est*, which can be compared with the above parameters.

	Image		
Overall Satisfaction	low	med	high
low	0.201	0.001	0.012
med	0.617	0.440	0.164
high	0.181	0.559	0.824

Table 7.3.26 Marginal conditional probabilities for Overall Satisfaction given Image for model est.

	ProdQuality		
Overall Satisfaction	low	med	high
low	0.051	0.012	0.114
med	0.607	0.345	0.323
high	0.341	0.643	0.563

Table 7.3.27 Marginal conditional probabilities for Overall Satisfaction given ProductQuality for model est.

	Price		
Overall Satisfaction	low	med	high
low	0.115	0.008	0.007
med	0.463	0.435	0.216
high	0.421	0.557	0.776

Table 7.3.28 Marginal conditional probabilities for Overall Satisfaction given Price for model est.

In contrast to the results on marginal conditionals for latent constructs given their respective features, we can conclude that the EM-ML estimation for OverallSatisfaction given service dimensions handled as latent constructs successfully recovers the conditional probabilities for OverallSatisfaction. This can be affected by the fact that OverallSatisfaction is an observed variable. At least, these observations can be made on the basis of the marginal conditional probabilities.

7.4. Determining categories of features

Because both satisfaction with service features and overall satisfaction have three states, in general we could represent each level of overall satisfaction as a function of probability corresponding to each level of satisfaction with feature. This would be a tedious and not useful task. Instead, as in the previous chapter, we take into account only two parameters that will be varied: $p(F_i$='high') and $p(F_i$='low'), where F_i is a satisfaction with feature i. On the level of overall satisfaction, we consider probability of high p_h(Sat), and low p_l(Sat) overall satisfaction.

In Tables 7.4.1 and 7.4.2 we show the probabilities of overall satisfaction needed to derive the coefficients for the model with observed dimensions. For instance, the first two rows in the Table 7.4.1 present the marginal probability that overall satisfaction level will be high as a function of the parameter $p(F_i$='high'). In the first row, we report the probability of high overall satisfaction p_h(Sat) given the parameter value is 0. This means that we set the prior probability of high satisfaction with the respective feature to 0, normalize the remaining two probabilities to sum up to 1, next we update the beliefs, and note the marginal probability of high overall satisfaction. Overall satisfaction is high with probability 0.533, if we set the prior probability of Im1 to 0, as an example. In the second row, we report the probability of high overall satisfaction p_h(Sat) given the parameter value is 1. Setting the parameter value to 1 is here equivalent to instantiating the feature nodes to high satisfaction. Consistently, high satisfaction with Im1 results in overall satisfaction with the probability of 0.585. Likewise, in Table 7.4.2, marginal probabilities of low overall satisfaction p_l(Sat) given different probabilities of low satisfaction with features $p(F_i$='low') are presented.

	Im1	Im3	Im4	PQ1	PQ3	PQ5	Pr1	Pr2	Pr5
p_h(Sat)(0)	0.533	0.542	0.548	0.541	0.535	0.544	0.543	0.542	0.541
p_h(Sat)(1)	0.585	0.555	0.525	0.567	0.568	0.547	0.571	0.600	0.575

Table 7.4.1 The probability ranges for sensitivity of high overall satisfaction for model *obs*.

	Im1	Im3	Im4	PQ1	PQ3	PQ5	Pr1	Pr2	Pr5
p_l(Sat)(0)	0.052	0.057	0.061	0.057	0.058	0.059	0.049	0.053	0.058
p_l(Sat)(1)	0.081	0.068	0.060	0.067	0.073	0.065	0.064	0.062	0.063

Table 7.4.2 The probability ranges for sensitivity of low overall satisfaction for model *obs*.

For the model with dimensions estimated from data we have found the following sensitivities presented in Tables 7.4.3-4. The tables should be read in a similar way as above.

	Im1	Im3	Im4	PQ1	PQ3	PQ5	Pr1	Pr2	Pr5
p_h(Sat)(0)	0.511	0.527	0.547	0.545	0.543	0.551	0.545	0.542	0.553
p_h(Sat)(1)	0.676	0.628	0.547	0.557	0.558	0.533	0.597	0.687	0.498

Table 7.4.3 The ranges for model sensitivity of high overall satisfaction for *est*.

	Im1	Im3	Im4	PQ1	PQ3	PQ5	Pr1	Pr2	Pr5
p_l(Sat)(0)	0.032	0.053	0.065	0.061	0.062	0.056	0.023	0.027	0.059
p_l(Sat)(1)	0.132	0.078	0.053	0.058	0.049	0.076	0.072	0.069	0.061

Table 7.4.4 The ranges for model sensitivity of low overall satisfaction for *est*.

Knowing that the relationship between p(Sat)(1) and p(Sat)(0) is linear, we can easily derive the values of the linear coefficients b_h and b_l for sensitivity of overall satisfaction as a function of satisfaction with a feature. Clearly, these values can be calculated as

$$b_h = p_h(\text{Sat})(1) - p_h(\text{Sat})(0),$$
$$b_l = p_l(\text{Sat})(1) - p_l(\text{Sat})(0).$$

Table 7.4.5 contains values obtained by performing this calculation. As can be seen, the coefficients' values b_h and b_l are generally lower in the absolute sense than the values found in the previous study. This is the case both for model *obs* and model *est*. This is because in the previous investigation, we have studied the classification of service dimensions, whereas here we deal with service features. The features have thus less influence than dimensions on the overall satisfaction, because in the current study the dimensions mediate the link between features and overall satisfaction, so the influence of features is absorbed by dimensions [see Madigan *et al.*, 1996].

	Im1	Im3	Im4	PQ1	PQ3	PQ5	Pr1	Pr2	Pr5
Model obs b_h	0.052	0.013	-0.023	0.026	0.033	0.003	0.027	0.057	0.034
Model est b_h	0.166	0.101	0.000	0.012	0.015	-0.019	0.052	0.145	-0.055
Model obs b_l	0.028	0.010	-0.001	0.010	0.015	0.006	0.015	0.009	0.004
Model est b_l	0.100	0.025	-0.012	-0.003	-0.013	0.020	0.049	0.043	0.002

Table 7.4.5 The coefficients b_h and b_l for both models.

7.4.1. Qualitative comparison

If we nonetheless accept the values of the coefficients b_h and b_l as significant enough then this assumption enables us to the classify the features. The definition of categories we use here is the same as in the previous study. We have taken the value t=0.01 as a threshold to determine the magnitude of influence, so that values of the coefficients b_h and b_l greater than 0.01 are considered as high, otherwise as low. Generally, we expect that these coefficients are positive real values. However, if the value is negative, then more insight is required to determine the category of the feature, and in such cases we put a label "undeterminable".

In Table 7.4.6, we show the category of each feature both for model with observed dimensions (Model *obs*), and for model with dimensions estimated from data (Model *est*).

Attribute		**Model *obs***	**Model *est***
Im1	Trust	exciter	satisfier/dissatisfier
Im3	Technology	non- relevant	exciter
Im4	Security	undeterminable	undeterminable
PQ1	Top technology	non- relevant	undeterminable
PQ3	User-friendliness	exciter	undeterminable
PQ5	Information	non- relevant	undeterminable
Pr1	Hardware prices	non- relevant	satisfier/dissatisfier
Pr2	Transaction prices	exciter	satisfier/dissatisfier
Pr5	Transparency prices	exciter	undeterminable

Table 7.4.6 Classification of features in model with three features per dimension.

Due to small values of the coefficients, and despite the lower threshold, we must conclude that we were unable to find any influence of four features on overall satisfaction with Model *obs*. These features (Im2, Pq1, PQ3, and Pr1) should therefore be classified as non-relevant. Four other features could be assigned the category "exciter."

For model with dimensions estimated from data by means of the ML optimisation, we have found that even more variables are not subject to any reliable classification. For instance, it turns out that the influence of low satisfaction with PQ2 on low overall satisfaction is negative, which would suggest that the more probability of low satisfaction that the any impact on overall satisfaction. Although also four features can be said as exceeding the threshold of 0.01, as much as the other five features do not lend themselves to classification.

Normally, the qualitative comparison involves tracing the different conclusions in terms of classification of the service features. When we accept the circumstances of low influence of features on overall satisfaction as reported in Table 7.4.6, we can conclude that the classification of features ends up with different results, which should be interpreted so that derivation of probabilistic parameters relating to service dimensions from data yields different local probabilistic distributions, at least for the data and features at hand.

In conclusion, we must state that the approach with introducing mediating variables is not feasible in practice because the influence of features on overall satisfaction turns out to be not significant. We cannot classify most of the them, and even for these features that, on the basis of Table 7.4.6, could be classified, the classification by the two models is different.

Another important finding is that our categorisation of features into satisfiers/dissatisfiers, exciters, basic, and non-relevant features might need to be revised and probably extended by new categories. This finding is due to possible negative relation between low levels of satisfaction with feature and low overall satisfaction, as well as high satisfaction with feature and high overall satisfaction. For more insight we could also consider models containing only relationships between overall satisfaction and one feature at a time. These models should be fully parameterised on the basis of all available data.

We should therefore conclude that measuring of the satisfaction on the level of the service dimension is necessary.

7.4.2. Two-step approach

Since according to our models, the influence of service features on overall satisfaction is small and cannot be significantly assessed, we propose a two-step approach for the purpose of gaining more insight into the nature of relation between features and the overall satisfaction. First, we classify the features with respect to their influence on service dimensions, and next we classify the dimensions with respect to their influence on overall satisfaction.

By performing analysis of the influence of features on service dimensions we can find out whether the local conditional probabilities estimated for model est enable classification of features with respect to their impact on satisfaction with service dimensions.

1. Influence of features on dimensions

Tables 7.4.7 and 7.4.8 present classification of features on dimensions.

	Im1	Im3	Im4	PQ1	PQ3	PQ5	Pr1	Pr2	Pr5
P(Sat)(0)	0,033	0,054	0,060	0,093	0,096	0,100	0,007	0,007	0,005
P(Sat)(1)	0,160	0,088	0,063	0,266	0,174	0,182	0,001	0,001	0,023

Table 7.4.7 The probability ranges for model *obs*.

	Im1	Im3	Im4	PQ1	PQ3	PQ5	Pr1	Pr2	Pr5
p(Sat)(0)	0,274	0,335	0,373	0,169	0,177	0,186	0,701	0,763	0,833
p(Sat)(1)	0,610	0,455	0,364	0,256	0,366	0,257	0,914	0,893	0,895

Table 7.4.8 The probability ranges for model *obs*.

For the model with dimensions estimated from data we have found the following sensitivities presented in Tables 7.4.9-10.

	Im1	Im3	Im4	PQ1	PQ3	PQ5	Pr1	Pr2	Pr5
p(Sat)(0)	0,303	0,285	0,335	0,337	0,407	0,317	0,251	0,238	0,266
p(Sat)(1)	0,528	0,618	0,449	0,675	0,343	0,632	0,348	0,736	0,180

Table 7.4.9 The ranges for model est.

	Im1	Im3	Im4	PQ1	PQ3	PQ5	Pr1	Pr2	Pr5
p(Sat)(0)	0,135	0,239	0,135	0,148	0,195	0,162	0,136	0,175	0,479
p(Sat)(1)	0,645	0,379	0,245	0,323	0,342	0,397	0,593	0,572	0,494

Table 7.4.10 The ranges for model est.

In Table 7.4.11, we show coefficients b_h and b_l for sensitivity functions of high and low satisfaction with service dimensions given probabilities of high and low probability with satisfaction with respective features.

	Im1	Im3	Im4	PQ1	PQ3	PQ5	Pr1	Pr2	Pr5
Model obs b_h	0,127	0,035	0,003	0,173	0,078	0,081	-0,006	-0,006	0,019
Model est b_h	0,225	0,333	0,114	0,337	-0,063	0,316	0,096	0,497	-0,086
Model obs b_l	0,337	0,120	-0,009	0,087	0,189	0,071	0,214	0,130	0,062
Model est b_l	0,510	0,140	0,110	0,175	0,147	0,235	0,457	0,398	0,015

Table 7.4.11 The coefficients b_h and b_l for both models derived from Tables 7.4.7-10.

It is easy to notice that the values of coefficients b_h and b_l are now high enough to categorize the features with their influence on dimensions. Consequently, Table 7.4.12 contains categories of the features with respect to their influence on dimensions. We have been able to classify some of the features, although most of them, namely Im3, PQ2, Pr1, Pr2, and Pr3 exhibited negative influence on general satisfaction. For Im3, Pr1 and Pr2 the absolute value of these coefficients are very low, so we assume that the size of the influence can be considered as low, and that these features can undergo classification. In such cases, we put a star next to the category, as shown in Table 7.4.12. As regards PQ2 and Pr3, these values are reasonably high, so we leave these features without assigning categories.

Attribute		**Model *obs***	**Model *est***
Im1	Trust	Satisfier/dissatisfier	Sat/ dissatisfier
Im3	Technology	Basic	Sat/ dissatisfier
Im4	Security	Non-relevant*	Non-relevant
PQ1	Top technology	Exciter	Sat/ dissatisfier
PQ3	User-friendliness	Basic	undeterminable
PQ5	Information	Non-relevant	Sat/ dissatisfier
Pr1	Hardware prices	Basic*	Basic
Pr2	Transaction prices	Basic*	Sat/ dissatisfier
Pr5	Transparency prices	Non-relevant	undeterminable

Table 7.4.12 Classification of features with respect to their influence on dimensions.

We notice that only five out of nine features can be classified in the same category by both models. Only trust, security, user-friendliness, hardware prices, and transparency of prices exhibit the same kind of influence on their relevant dimensions. The same nature can be found only by accepting that two out of these five features exhibit minimally negative influence on satisfaction with service dimensions.

We must conclude that in the light of this result, it makes little use to compare the models by means of the effects of service dimensions on overall satisfaction.

7.4.3. Conclusions on the mediated model

Firstly, the mediated model of overall satisfaction with service dimensions for which data are observed does not allow for reliable classification of features because of the small derived effect of features on general satisfaction. Therefore, possible non-mediated models, in which features are directly linked to general satisfaction, should be therefore proposed and tested.

Moreover, we found that the nature of the relationships between satisfaction with service attributes and service dimension on the one hand, and between service dimensions and overall satisfaction on the other hand is complex and not straightforward. Some features exhibit unexpected impact on service dimensions. We acknowledge that some of the items may be regarded as lacking face validity, therefore this result could be justified.

Therefore, due to this complex nature, estimation of hidden nodes in CS research labelling by means of ML can be cumbersome. Aliasing is a crucial problem in this context. It can be tackled by assigning non-uniform prior information to the hidden constructs, and doing the MAP optimisation rather than the ML optimisation. By assigning non-uniform prior distribution we impose states on the values of hidden variables. Furthermore, we speculate that if the

CPT tables were smaller, allowing for better use of observed data, then ML estimation could be more successful.

It is not necessary to measure customer satisfaction at the dimension level when the main purpose of CS study by means of Bayesian networks is to predict the level of overall satisfaction.

However, when we are concerned specifically with the feature importance-performance analysis, we have found that derivation of probabilities relating to service dimensions only on the basis of data on service dimensions and overall satisfaction has failed. Moreover, the learned probabilities do not allow for the use of the model in the feature importance-performance analysis.

7.4.3.1. Implications

The most important implication for client-oriented businesses resulting from this work is that measuring satisfaction at the level of service dimensions is necessary.

Possible non-mediated models in which features are directly linked to general satisfaction should be therefore proposed and tested.

7.5. Classification of features with noisy OR-gate

In the previous study, we have found among others that the mediated model is not successful because the effect of features on overall satisfaction is to a great extent absorbed by the mediated layer, i.e., satisfaction with dimensions. In order to overcome this problem, in this study we apply and evaluate a non-mediated model of overall satisfaction in which service features are direct parents of overall satisfaction.

Another modelling technique to decrease complexity of the models like shown in Figure 7.1.1 is introducing the noisy OR-gate [Good, 1961; Pearl, 1988]. This type of dependence is especially suitable when one deals explicitly with multiple causes and one effect. Another name, under which this form of dependency is known, is *disjunctive interaction*, because the causes act disjunctively.

The primary objective of modelling overall satisfaction with the noisy OR-gate is to test whether it is suitable for a specific classification of service features as winners, or must-be or penalty attributes.

7.5.1. Definition of noisy OR-gate

The noisy OR-gate makes three assumptions on the causes end effects. These assumptions are causal inhibition, accountability, and exception independence [Neapolitan, 2004].

Firstly, the concept of noisy OR-gate is based on the assumption of *causal inhibition*. It implies that there exists some mechanism that prevents, or inhibits, a cause from bringing about its effect, so that this impact of the cause on effect is fully effective only if the mechanism that prevents this impact is inactive.

Secondly, *accountability* entails that an effect takes place when at least one of its causes is present and is not inhibited. Although this might seem very strict at the first sight, it follows that all causes which are not explicitly taken into account must be put together into one unknown cause that can be labelled "All other causes". This cause can be also viewed as a *background* cause that is always active.

Thirdly, *exception independence* entails that the processes that inhibit one cause are independent from the processes that inhibit all other causes.

The above-mentioned assumptions can be portrayed schematically as in Figure 7.5.1. This figure shows the situation in which there are n causes X_1, X_2, ..., X_n of Y, each heaving some prior probability of occurrence. The variable I_j is the mechanism that inhibits X_j. These I_j variables are independent due to the assumption of exception independence. Next, the variable A_j is on, if and only if X_j is present (true), and the mechanism that inhibits X_j is not active. Owing to the assumption of causal inhibition, this means that Y should be present if any one of the A_j's are present. The dependency between Y and A_j can be seen as a logical or, in the sense that if at least one of A_j is present (true) then Y is also present (true), otherwise it is absent (false). That is why the model is called an "OR-gate".

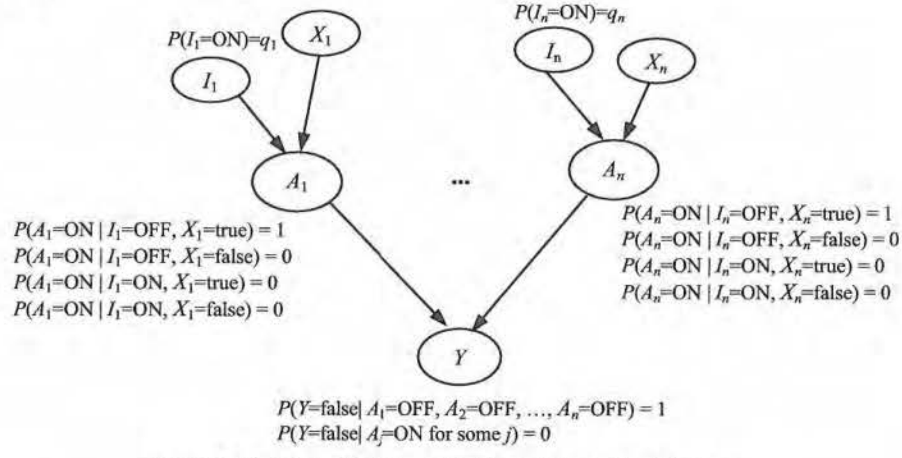


Figure 7.5.1 Schematic representation of the noisy OR-gate.

We will denote by q_j the probability that the jth inhibitor is active. If X_j is the only parent that is true, Y will be true if and only if the inhibitor associated with X_j remains inactive. We have

$$P(Y = false \mid X_j = true, X_i = false \text{ for } i \neq j) = q_j.$$

Now we will derive a closed-from calculation of the probability distribution of Y given any assignment of w [Pearl, 1988]. Let W = $\{X_1, X_2, ..., X_n\}$, and let w = $\{x_1, x_2, ..., x_n\}$ be a set of values of variables in W. Furthermore, let S be a set of indices such that $j \in$ S if and only if $X_j = true$. That is, S = $\{j$ such that $X_j = true\}$. Then

$$P(Y = false \mid \mathrm{W} = \mathrm{w}) = \prod_{j \in \mathrm{S}} q_j \tag{7.1}$$

Equation 7.1 is the basis for the probabilistic inference in Bayesian networks with noisy OR-gate, but more efficient algorithms exist that incorporate the noisy OR-gate model in the inference algorithms, such as Pearl's message passing algorithm [Pearl, 1988]. Alternatively, we can express the probability that Y is true as

$$P(Y = true \mid \mathrm{W} = \mathrm{w}) = 1 - \prod_{j \in \mathrm{S}} q_j \tag{7.2}$$

Often, it is more convenient to specify p_j as

$$p_j = 1 - q_j, \tag{7.3}$$

which is called the *causal strength* of X_j for Y. From Equation 7.2 it follows that

$$p_j = P(Y = true \mid X_j = true, X_i = false \text{ for } i \neq j). \tag{7.4}$$

The probability p_j can often be much more accessible than conditional probabilities assuming the conditional multinomial distribution. It can be also relatively easy estimated from the data. Assume we have a database of cases. For example, if we want to estimate the causal strength of any cause we would count the number of cases in which the cause is the only active cause, and how many of the cases with only this cause active have also the effect variable active.

The most important consequence of the noisy OR model applied to the service feature satisfaction study is that the number of probabilities needed to estimate the model grows linearly with the number of features. This comes unfortunately with the restrictive assumption of non-interactivity between features, i.e., the configurations of parents' states, in which more than one parent are active, are ignored during parameterisation of the noisy-Or gate. In consequence, the noisy-Or gate is less expressive than the full conditional probability table.

7.5.2. Overall satisfaction as the noisy OR-gate

Let us assume that each service feature acts as a potential direct causal factor for overall satisfaction in line with the model in Figure 7.1.1. Let us focus our attention on high and low levels of overall satisfaction, in which case we will refer to as overall satisfaction and dissatisfaction, respectively. Furthermore, we will assume that service features can also take only two states: if satisfaction with the feature is high will say that the feature is present, or active; if satisfaction is low, then the feature is said to be absent.

By considering overall satisfaction as the noisy OR-gate, we presume firstly that the more features are active, the higher the chance for satisfaction feelings with the service/product in general – a reasonable assumption. Secondly, satisfaction judgments with different features are assumed to be independent. In our opinion, this assumption can be more justified for features relating to different service dimensions, and less realistic with respect to features within the same dimension. Thirdly, and more importantly, the disjunctive interaction implies that we exclude *a priori* the plausible interaction effects (not only negation but also synergy) between service features. This assumption could be too restrictive in practice, but since we are not focused here on probabilistic simulations or inference, we have decided to try out this modelling task without including the interaction effects in this analysis.

Let us see how the other assumptions specific to the noisy OR-gate can be translated into the context of customer satisfaction modelling. Assumption of causal inhibition requires that even if a customer is satisfied with a service feature, then there is a chance that this satisfaction does not bring about overall satisfaction, since a mechanism exists that may prevent it. Exception independence requires that these mechanisms that prevent the impact of satisfaction with different features on overall satisfaction are independent. Accountability implies that a customer cannot be overall satisfied unless at least one of service features, including features collected under the label "other features", is satisfying.[1] In our opinion, all the three assumptions can be seen as fulfilled.

We assume here that modelling with noisy OR-gate can be appropriate for finding features that we can call winners on the one hand, and penalty (or must-be) features on the other hand. Winners can be defined as those satisfactory features that, whereas other features are less than satisfactory, still have the effect that overall satisfaction is high. Conversely, a service attribute that as the only dissatisfying one makes that satisfaction on the whole is also very low can be deemed as a penalty, or a "must-be" attribute.

7.5.3. Model specification

In order to apply the noisy OR-dependency to our data, we had to recode the responses so that the variables were binary. Recall from the section on measurement that overall satisfaction with measured on a 10-point rating scale ranging from 1 to 10, whereas satisfaction with features was measured on a 5-point scale ranging from 1 to 5. The response valued 1 was meant as "completely disagree", or "very dissatisfactory". We have assumed that the responses on satisfaction with features higher than 4 reflect the situation in which high satisfaction is present. This means that the values "strongly agree" have received the labels "satisfaction". As regards overall satisfaction, the same procedure was applied with the equivalent threshold of 7. In order to assess influence of dissatisfaction with features on overall dissatisfaction, we recoded the responses so that the score lower than 4 on features was regarded as dissatisfaction with the feature, and the score lower than 6 on overall satisfaction was regarded as overall dissatisfaction.

[1] Accordingly, if we were concerned with the noisy OR-model of dissatisfaction, we would speak about dissatisfactory judgments of service features.

In order to determine importance of features for overall satisfaction, we have constructed several models, each with features of another service dimension as parents of overall satisfaction. Because by definition, less parameters are required to parameterised the model, we have decided to take more features, namely four, as parents of overall satisfaction into account than we did in the previous section. In fact, we have performed the similar procedure in which we have let the number of features vary between three and six. We have found that reliable parameterisation is relatively difficult to achieve.

Figure 7.5.2 shows the structure of this model on the example of Image features, taking into account four features Trust (Im1), Technology (Im3), Competitiveness (Im6), and Professionalism (Im8). Besides Image, we have also performed the analysis with two other service dimensions, namely Product Quality and Prices. As regards Product Quality, the set contained the following randomly selected features: Top technology (PQ1), Durability (PQ2), User-friendliness (PQ3), and Information (PQ5). For Prices, Honesty (Pr1), Transaction prices (Pr2), Transparency (Pr5), and Customisation (Pr6) were selected. Although it is not shown in this figure, the conditional probabilities table for Overall Satisfaction is described by the noisy-OR function. The tables for features contain prior probabilities estimated from data.

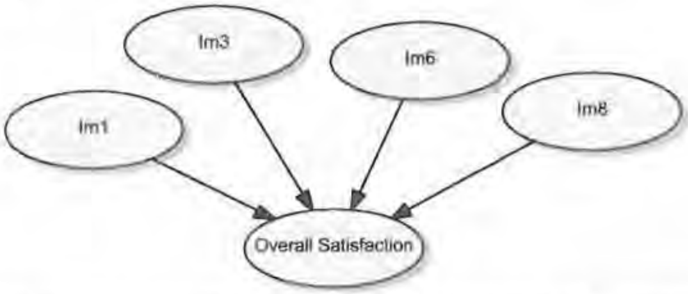


Figure 7.5.2 Noisy-OR model of customer satisfaction with Image features. Dependency of Overall Satisfaction and the features is in fact described by the noisy-OR distribution.

All the probabilities, including those needed to specify the parameters of the noisy-OR function, were estimated on the subset of the original data, which contained only those cases for which no data was missing on all the modelled variables.

7.5.4. Results

Let us first take a look at the values of the estimated parameters of the models. We will explain the results on the example of features related to Image, and later we will address features of other dimensions.

Image

For impact of Image features, we obtained the results shown in Table 7.5.1. This table has two parts. In the upper part, in the row named "High", we show influence of high satisfaction with features on high overall satisfaction. In the lower part, named "Low', the effects of low satisfaction with features on low overall satisfaction are shown. For the reason of clarity, we will focus on the upper part of the table, and discuss effects of high satisfaction with features on high overall satisfaction. Later, we will address the impact of dissatisfaction with features on low overall satisfaction (or dissatisfaction).

		Im1	Im3	Im6	Im8	Other
High	Causal strength	0.792	0.473	0.438	0.5	0.459
	Coverage	53	38	16	80	597
	Support	0.041	0.011	0.007	0.039	0.269
	Lift	1.727	1.032	0.953	1.089	1
Low	Causal strength	0.158	0.051	0.073	0.062	0.034
	Coverage	38	39	**220**	16	323
	Support	0.006	0.002	0.016	0.001	0.011
	Lift	4.636	1.506	2.135	1.835	1

Table 7.5.1 Causal strength of satisfaction with Image features on overall satisfaction (N=1020).

The influence of features on overall satisfaction is expressed with the parameter p_j reflecting the causal strength of the feature F_j. To be more precise, causal strength is the probability that overall customer satisfaction will be high given the respective feature is the only satisfying one among the four features shown. It can also be seen as the confidence of the association between the feature and the general satisfaction. On the other hand, it is also the probability that the inhibitor is not active.

According to the noisy-OR model there can always some level of high overall satisfaction be present when satisfaction with all the explicitly listed features is absent. So, in the column "Other", we show the configuration when none of the four listed features is satisfying, yet the overall satisfaction takes place.

The values in Table 7.5.1 should be interpreted as follows. For a specific feature F_j in focus, it would be desirable to observe high values of the causal strength parameter p_j, since high values would indicate that there is significantly higher probability of satisfaction than dissatisfaction if the feature is the only one satisfying feature. Of course, the support should be high as well, since only then can we have confidence in the value of the causal strength. As regards the column named "Other", it would be desired if the causal strength, call it p_O, were possibly low. When a customer is not very satisfied with all the explicitly considered features, we would expect that he/she tends to be dissatisfied rather than satisfied with the service in general. If this is the case, then the features that we explicitly account for in the table can be viewed collectively as most important features; alternatively, if the causal strength of "other" features is too high, then it provides an indication that the features we chose explicitly are not relevant for overall satisfaction.

Coverage is the number of respondents that are satisfied only with the feature in focus, and, at the same time, dissatisfied with all other features referring to the dimension in focus, regardless of the level of overall satisfaction. It is therefore an important measure of prevalence of the pattern, and should be thus ideally very high. Furthermore, we show also support of the association, which we define as the ratio of number of respondents who are satisfied both overall and with one specific feature in focus by the number of all respondents for which all the relevant variables are observed. Support should also have high values, however, preferably, support of all other features should be minimised since then we would hopefully be able to explain more overall satisfaction with features selected explicitly.[1]

For more insight into the effect of satisfaction with particular features on overall satisfaction, we should relate the causal strength of the feature to the causal strength of all other features not mentioned explicitly by dividing p_j by p_O. We can call the resulting score as *lift* and define it as a measure how much more superior the feature is compared to all other features in relation to overall satisfaction. By means of lift, we can also identify the category of each feature. High lift with respect to high overall satisfaction indicates that the attribute is a winner, while high lift with respect to low overall satisfaction indicates that the attribute is a penalty attribute. Lift is shown in the last row in each part of the table.

Let us now analyse the most interesting results for features related to Image. For instance, we can observe that high satisfaction of feature Im1 (in this case it is Trust) has the causal strength of 0.792 on high overall satisfaction, which means that the probability of high satisfaction, given the only satisfying feature of the three considered is Im1, is 0.792. In other words, there is 0.208 probability that the inhibitor, that suppresses the influence of Im1 on overall satisfaction, is active. The causal strength of Trust is thus relatively high. Coverage is 53 and support amounts to 0.041, so in our opinion these two values are rather low, and do not guarantee that the causal strength is significant for the entire population of respondents. Second strongest causal impact, according to our noisy-OR model, is exhibited by Professionalism. Coverage for the two other features is too low to draw conclusions. As can be seen, there is probability of 0.459 that the customer is overall contented due to satisfaction with other "background" features.

For influence of low satisfaction with features on low general satisfaction, it can be observed that very high lift is recorded for Im1, but coverage is not high

[1] We could therefore perform a feature selection procedure so as to minimize the causal strength of "other" features, but in order to keep the discussion clear we refrain from doing that.

enough. Moreover, we notice that Im6 (Competitiveness) is very often the only dissatisfying feature, as coverage equals 220. This number is big enough for considering the pattern as reliable. However, only 7.3% is at the same time dissatisfied overall, whereas the rest of respondents, i.e., 92.7%, is moderately or highly satisfied. It shows that this feature has in absolute values, on its own, very low impact on dissatisfaction judgments in general with the electronic payments services. However, when we take into account causal strength of features collected in the columns "other", then, relatively, this feature has a 2.135 times bigger impact than these other features. Due to this value, we can conclude that Competitiveness of the company in the market (Im6) is a must-be (penalty) feature. Consequently, the company must have a competitive position in the market, since customer dissatisfaction in this respect can very likely go together with overall dissatisfaction with the service.

Price

Similar feature importance analysis can be performed for selected features of Price. The results are shown in Tables 7.5.2. For determinants of low overall satisfaction, when dichotomising, we have adjusted the threshold from 4 to 3 as there were too few respondents that were highly satisfied with the features.

		Pr1	Pr2	Pr5	Pr6	Other
High	Causal strength	0.5	0.625	0.609	0.643	0.519
	Coverage	12	8	64	14	863
	Support	0.006	0.005	0.039	0.009	0.444
	Lift	0.963	1.204	1.174	1.238	1
Low	Causal strength	0.076	0.121	0	0.090	0.015
	Coverage	26	33	15	55	256
	Support	0.002	0.004	0	0.005	0.004
	Lift	4.92	7.757	0	5.818	1

Table 7.5.2 Causal strength of satisfaction with price features on overall satisfaction (N=1008).

The configurations in which the selected features of Price are the only active parents of overall (high) satisfaction are too seldom. In total, the sum of the specific coverage amounts to 98, out of 1008 cases. Similarly, low coverage and support can be observed for associations between attributes and overall dissatisfaction (low satisfaction). Consequently, we must note that none of the Price features can be successfully classified with the applied Noisy-OR approach.

Product Quality

Finally, we take a look at the selected features of Product Quality and their effect on overall satisfaction. The results can be found in Table 7.5.3. On the basis of low total coverage of the four features, equal to 126, it can be implied that we cannot draw reliable conclusions about their influence on high overall satisfaction. With respect to dissatisfaction, the only service attribute with high coverage, as of 129 out of total 870 cases, of its association with overall dissatisfaction is Top Technology of transaction terminals (PQ1). This association is also relatively strong, with lift of 2.967. As a result, we can regard the Top Technology of products as a must-be feature. In other words, the technology of products must be as high as possible; otherwise, there is a high chance that customers dissatisfied with this feature will be also dissatisfied overall with the entire service.

		PQ1	PQ2	PQ3	PQ5	Other
High	Causal strength	0.556	0.476	0.589	0.526	0.513
	Coverage	9	42	56	19	520
	Support	0.006	0.023	0.038	0.011	0.307
	Lift	1.082	0.927	1.148	1.025	1
Low	Causal strength	0.194	0.133	0.071	0.111	0.065
	Coverage	**129**	30	14	36	444
	Support	0.028	0.004	0.001	0.004	0.033
	Lift	2.967	2.041	1.093	1.701	1

Table 7.5.3 Causal strength of satisfaction with product quality features on overall satisfaction (N=870).

7.5.5. Conclusions on the Noisy-OR model

Prior to the analysis, we have expected that the main advantage of the noisy-OR model would be that less parameters for the dependency between the overall satisfaction and service features are needed to be estimated. As a consequence, administrators of customer satisfaction studies willing to use the noisy-OR model might require engaging fewer respondents in a study, which would be a very positive aspect for practical satisfaction studies.

Because of the non-interactive nature of the influence of features on overall satisfaction and its parameterisation, we have expected that modelling with the noisy OR-gate could be expected to be best suitable for detecting "must-be" or "winner" features. However, disjunctive interaction implies also that we consider only those cases in the data for which responses were in some specific configuration. In practice, we have found that many records cannot be used in parameterisation of the noisy OR-model since too many respondents are satisfied with more than one feature at a time. These cases are actually abandoned in the model parameterisation, which has in turn negative consequences for reliability of parameters. It has turned out that the number of cases in this study was too small to yield reliable patterns on the importance of service features.

In general, two conditions must be fulfilled to successfully determine the category of the feature. Firstly and most obviously, the causal strength must be high. Secondly, we can perform classification of features only if coverage is high enough.

We have found that none of the twelve features included in this study when acting on its own has significant effect on high general satisfaction with the service. Furthermore, we have found that in the case of only two features out of twelve considered, these two conditions are satisfactorily fulfilled. More specifically, we found that Top Technology (PQ1) of products is of great importance in contributing to overall dissatisfaction, and Competitive position of the company in the market (Im6) alike.

Therefore, we conclude that the Noisy-OR model of overall satisfaction have turned out useful only in detection of "must-be" features. We have not found other categories of features than "must-be".

7.5.5.1. Implications

The research we have carried by means of noisy OR-gate provides some useful knowledge for managers.

Firstly, the terminals must be technologically on the top, as TopTechnology is a must-be feature. This means that when customers have favourable experience with other features, and at the same perceive the technology of electronic payments terminals as outdated then the chances that they are overall dissatisfied is much bigger than on average. Secondly, Competitiveness of the company plays an important role, and can also be regarded as a penalty factor.

Furthermore, we can draw also important implications for marketing modelling. Future models should make use of all combinations of features' values in the data and thus allow for interaction effects. One reason for this is that the data in which one excludes the interaction effects become scarce and do not allow for reliable parametric estimation.

Therefore, another possibilities of modelling dependencies should be explored allowing for interaction effects among service features. From among the techniques worth consideration, logit models of order higher than one should be considered. Higher order models enable studying of interaction effects and improve representational expressiveness of the dependencies. In the future, we suggest considering second- and third-order models of interaction between features. However, studying models with order higher than 3 is interpretationally complex and runs the risk of unreliable estimation of necessary parameters. Other alternative models include generalizations of the noisy OR-gate, such as the noisy max-gate, and other models that allow for interactivity among features [see Takikawa and D'Ambrosio, 1999].

7.6. Conclusions and future research

7.6.1. Conclusions

In this case study, the predominant objective was adapting and evaluating Bayesian network for classification of features.

a. More specifically, we evaluated the mediated model of overall satisfaction based on the technique of parent divorcing in the analysis of feature importance. In the mediated model that we have evaluated, customer satisfaction at the dimension level mediates the link between service features and overall satisfaction. The results indicate that the investigated approach does not perform successfully. To be precise, we found that such a model does not allow for reliable classification of features because of the small derived effect of features on general satisfaction.

We found also that the classification is not feasible because the relationships between service features and dimensions are too complex for the proposed Bayesian network technique based on sensitivity analysis. In this respect, too many features manifest negative influence on overall satisfaction. For instance, it turns out that for some features, the higher satisfaction with that feature, there is less probability of high satisfaction with a dimension. These observations call for more investigation. Possible explanation is that respondents tend to classify the dimensions and features in different ways.

b. Secondly, we pursued to find out whether in the mediated model it is possible to treat satisfaction with service dimension as a hidden node, and thus optimise a questionnaire by not asking about satisfaction with service dimension. Leaving out these questions from the questionnaire would be advantageous for customer-oriented companies, since it would simplify the measurement procedure, positively influence the reliability of the included concepts, and reduce the costs of the satisfaction programs.

We have examined models with two different parameterisations. One of them has been fully parameterised on the basis of observed data. In the other model we treated satisfaction with service dimensions as hidden variables, and we used the EM algorithm to estimate the necessary probabilistic distributions. We found that classification of features is different in each case. The reason for that is that the maximum likelihood estimates of the conditional probability tables estimated with the EM algorithm in the model with hidden service dimensions disagree with the respective conditional probabilities in the model with observed service dimensions. However, taking into account the predictive accuracy both models perform equally well.

Due to this complex nature, estimation of hidden nodes and labelling by means of Maximum Likelihood can be cumbersome. Aliasing is a crucial problem in this context, since we found it very difficult to retrieve the correct labels for satisfaction at the dimension level.

Interestingly, we have found that it is not necessary to measure customer satisfaction at the dimension level when the main purpose of CS study by means of Bayesian networks is to predict the level of overall satisfaction.

However, when we are concerned specifically with the feature importance-performance analysis, we have found that derivation of probabilities relating to service dimensions only on the basis of data on service dimensions and overall satisfaction has failed. As a result, the learned probabilities do not allow for the use of the examined model in the feature importance-performance analysis.

c. Our last objective was to evaluate the noisy-OR model of overall satisfaction in the analysis of feature importance in which features are direct parents of overall satisfaction.

Because of the non-interactive nature of the influence of features on overall satisfaction and its parameterisation, we have expected that modelling with the noisy OR-gate could be expected to be best suitable for detecting "must-be" or "winner" features. However, disjunctive interaction implies also that we consider only those cases in the data for which responses were in some specific configuration. In practice, we have found that many records cannot be used in parameterisation of the noisy OR-model since too many respondents are satisfied with more than one feature at a time. These cases are actually abandoned in the model parameterisation, which has in turn negative consequences for reliability of parameters. It has turned out that the number of cases in this study was too small to yield reliable patterns on the importance of service features.

In general, two conditions must be fulfilled to successfully determine the category of the feature. Firstly and most obviously, the causal strength must be high. Secondly, we can perform classification of features only if coverage is high enough.

We have found that none of the twelve features included in this study when acting on its own has significant effect on high general satisfaction with the service. Furthermore, we have found that in the case of only two features out of twelve considered, these two conditions are satisfactorily fulfilled. More specifically, we found that Top Technology (PQ1) of products is of great importance in contributing to overall dissatisfaction, and Competitive position of the company in the market (Im6) alike.

Therefore, we conclude that the Noisy-OR model of overall satisfaction have turned out useful only in detection of "must-be" features, as we have not found other categories of features than "must-be" in a reliable way.

7.6.2. Implications

We were able to find a couple of implications of this research for applied CS data analysts.

Indirect derivation of satisfaction with service dimensions entirely on the basis of response data on overall satisfaction score, and satisfaction with features turns out not to be successful. The nature of the relationships between performance of service features and the judgments of overall satisfaction is probably too complex for the ML optimisation to approximate the original relationships between features and overall satisfaction.

The finding that the categories of two models, one with observed dimensions and the other one with dimensions estimated, are different implies that we should operationalise also satisfaction at the dimension level in future customer satisfaction studies.

However, on the other hand, the results suggest that asking directly for some compound construct is not necessary if the main model's use is the prediction of overall satisfaction, since the same predictive power of the model can be obtained when the satisfaction with dimensions is treated as hidden. Otherwise, if the model is intended to be used in feature importance/performance study, it must be used with a lot of caution.

The model considered here can be used in a more complex Bayesian network model that involves other loyalty variables.

7.6.3. Limitations

The work presented in this chapter has a number of limitations.

The main limitation of both presented methodologies is that they still allow studying a very limited number of service features at a time in a feature performance/importance study. Since the conditional probability table is growing very fast with the number of features, and yielding nuisance with the model's parametric estimation. In order to include a small numbers of features we had to select features randomly.

With respect to parent divorcing we should take into account that the three features that we randomly selected may not be valid and reliable determinants of satisfaction with their respective service dimension. Therefore, it is not possible to recover reliably the conditionals $p(D_i| F_1, F_2, \ldots, F_6)$. The assumptions required for legitimate use of divorcing are probably violated.

We assumed that satisfaction judgments with respect to different service features are marginally independent with each other. In practice, they can be interdependent and related with each other.

In addition, with respect to noisy OR-model, a severe limitation of modelling overall satisfaction with this model is that this model assumes non-interaction effects among service features.

Also, since each feature was measured only with one item we are not able to test the reliability of the instrument, and it may follow that the meaning of an item is for one customer different than for another one. In consequence, it might also happen that the measures are not reliable, i.e., they do not measure the true construct under investigation [e.g., Bagozzi, 1994a].

7.6.4. Future research

The case study in this chapter opens several potential avenues for future research.

We found that aliasing poses a serious problem to successful classification of features in the mediated model. We speculate that this problem can be tackled by assigning non-uniform prior information to the hidden constructs, and doing the Maximum A Posteriori (MAP) optimisation rather than the Maximum Likelihood (ML) optimisation. By assigning non-uniform prior distribution we impose states on the values of hidden variables. Furthermore, we speculate that if the CPT tables were smaller, allowing for better use of observed data, then Maximum Likelihood estimation could be more successful. The examination of this assumption can be addressed in future work.

Another possibilities of modelling dependencies should be explored allowing for interaction effects among service features. From among the techniques worth consideration, logit models of order higher than one should be considered. Higher order models enable studying of interaction effects and improve representational expressiveness of the dependencies. In the future, we suggest considering second- and third-order models of interaction between features. However, studying models with order higher than 3 is interpretationally complex and runs the risk of unreliable estimation of necessary parameters. Other functional dependencies enabling interaction effects should be also considered, including generalizations of the noisy OR-model, such as noisy-MAX, noisy-MIN [Srinivas, 1993; Takikawa and D'Ambrosio, 1999].

Another issue worth investigation is to analyse the sensitivity of classification both with the mediated model as well as with the noisy-OR gate to different thresholds used to recode the original data.

8. Conclusions, limitations and future research

In this chapter, in Section 8.1 we collect and discuss the most important part of this manuscript – the final conclusions. Then, in Section 8.2, we discuss the implications of our work for CS&L research and managerial practice. Limitations of this research are addressed in Section 8.3, and we close with potential avenues for further research in Section 8.4.

8.1. Conclusions

In discussing the conclusions, we will first review the objectives related to the use of Bayesian networks in theoretical CS&L research, then in the practical CS research, and finally, we will review the strengths and weaknesses of Bayesian networks in general. We will discuss each objective at a time, referring to each of them as we did in Chapter 1, except for the last question in each part, concerning the strengths and weaknesses of Bayesian networks; the latter will be discussed collectively under one research question. Each research question is discussed in terms of its sub-objectives, followed by a summary.

<u>I. Conclusions with respect to the use of Bayesian networks in theoretical Customer Satisfaction and Loyalty research</u>

Our contribution to the marketing modelling literature in the first part of this work was to thoroughly examine the Bayesian network-based approach to theoretical modelling in the customer e-satisfaction and loyalty research. In this respect, we put three essential questions at the beginning of this thesis.

<u>1. How can marketing theories be discovered (developed) by means of the Bayesian network approach?</u>

1.a. In general, marketing theories can be developed taking the inductive or deductive approach.

1.a.i. In Chapter 4, we critically examined Bayesian networks in the inductive research. Since customer e-loyalty is a relatively new phenomenon, we decided to design an inductive study, in which on the basis of data we aimed to discover a possible theory of this phenomenon; we departed from a position in which we were not sure what could be the relationships between theoretical constructs in the domain; what we did assume was only the ordering of variables, from the most antecedent constructs to the ultimate ones. Next, we have let a search algorithm look for the most likely model given the data in the space of different theoretical hypotheses; since we had four different data sets for visitors of different web sites, this search was performed independently for all of them. The result was plausible: the learned models are very similar to each other in terms of theoretical consequences. We can thus observe that the inductive search with the Bayesian network approach makes reliable inference from data, at least when we compare the outcomes with the existing state of knowledge present in the e-loyalty literature. Hence, we conclude that the results obtained are generic, in the sense that the differences that exist in all possible aspects of each portal site considered, and most importantly web users' perception thereof, do not have any influence on the underlying theoretical model of e-loyalty. This finding suggests also that there exists an overall model of e-loyalty that is valid generally for portal web sites.

Therefore, based on this last finding, we proposed and carried out a procedure constructing an overall model of e-loyalty, derived from these four single models. Interestingly, we found this overall model also theoretically sound given the existing e-loyalty literature. Unfortunately, since the e-loyalty literature is relatively scarce to date, we were not able to affirm whether the overall model was fully confirmed by this literature. Consequently, our potential contribution to the e-loyalty phenomenon could be that the overall attitude towards a portal site is not so much important compared to the site quality, and especially the ease of navigation.

From the Bayesian network modelling perspective, we must conclude that not only the greedy nature of the algorithm that searches for the most likely model, but also the marginal likelihood score itself, as a measure of goodness of fit, proved very appropriate and successful in developing theory of customer e-loyalty.

It is necessary to note here that the inductive approach we have taken in this study is in fact very close to the exploratory research in that we can easily develop a theoretical model from cross-sectional response data without imposing any prior hypotheses that could bias the resulting theory; of course prior ordering of variables could be argued to be one of such prior information, therefore we discuss it in more detail in the section on limitations. In our opinion, we can nevertheless conclude that the Bayesian network approach is suitable for exploratory research.

All in all, we conclude that the performance of the Bayesian network approach in inductive e-loyalty research is successful and its examination is positive.

1.a.ii. In the deductive approach examined in Chapter 5, the researcher's goal is to empirically validate a theory that is a priori presumed. In this context, we advanced several competing hypothetical models of the CS&L theory on the basis of the literature. Our aim was to discriminate between these models. We found that taking the Bayesian score as a measure of the goodness of fit, we can validate a hypothesis of presence or absence of a direct relationship between two constructs.

In particular, from the results in Chapter 5, we can see that both the Cheeseman-Stutz and the Bayesian Information Criterion scores are highest for the model that suggests existence of a direct dependence of Involvement on Satisfaction and Trust. This dependence is more probable than the dependence of Involvement on Satisfaction only. Interestingly, the *a posteriori* most likely theoretical model has been, in its full form, postulated in fact by the marketing research company before seeing any data.

In conclusion on the deductive research with Bayesian networks, we argue that the deductive approach can be successfully carried out within the Bayesian network modelling; it must be remembered however that, unlike it is the case with other techniques applied in CS&L research, it is not possible with the presented approach to perform validation based on the marginal likelihood to strictly confirm whether the model can be accepted or should be rejected.

1.b. Correspondence between latent constructs and their measures has to be an explicit component of marketing models in CS&L research [e.g., Steenkamp and Baumgartner, 2000; Bagozzi, 1984]. Nowadays, no method exists, to our knowledge, of incorporating the structural and measurement models explicitly into the Bayesian network modelling. Therefore, we developed and evaluated a new method for handling structural and measurement models with the Bayesian network approach.

1.b.i. Let us first discuss our proposed method for incorporating the latent construct and measurement model in Bayesian network modelling. Our method of accounting for these two models in one holistic analysis, based on local Naïve Bayes models for modelling the dependency between a latent construct and its measurement items, can be seen as a contribution of a great importance. First of all, the results of our proposed method are theoretically sound in the sense that structural model that before seeing any data is assumed most likely, indeed scores best. It is apparent that the proposed method of handling the structural model can be used to test the presence or absence of some theoretical relationships between latent constructs. Furthermore, conditional probabilities between latent constructs, i.e., defining the structural model, are meaningful and provide valuable insight into the nature of relationships. Additionally, the relationships in the measurement model are also meaningful and show that the approach, which we proposed in this chapter, is valuable and performs well, at least when considered apart from any other, standard techniques. The proposed measurement modelling approach has however a major drawback of not being able to control qualitatively for the measurement error.

Furthermore, we have performed comparison with a straightforward approach in which latent constructs are not treated as latent but are constructed as the average over indicator variables. We have found that our proposed latent construct model outperforms this standard approach in the classification accuracy. Taking the average is not the optimal technique probably because it ignores the relative importance of the indicator variables in measuring the abstract concept. So, there is potential loss of valuable information. This loss can also be seen by interpreting the classification accuracy, where averaging of indicators for Trust and Involvement results in worse classification function for Loyalty. Secondly, there exist only slight differences in conditional distributions. Conditional distributions are sharper in case of the "averaged items" approach, whereas for the latent construct model they tend to be softer and more alike the uniform distributions. Thirdly, unlike the "averaged items" approach, the latent construct model allows for assessing the validity of the scale.

In summary on our proposed latent construct modelling approach by means of local Naïve Bayes, we conclude that our approach performs well. Our method proves to be useful and shows the added value of this work.

1.b.ii. In addition, we proposed a technique of latent construct validation in Bayesian network modelling. To be precise, the proposed method enables validation of the measurement instrument to the extent that the effect indicators are related either to one latent construct or to two potential latent constructs. In other words, it enables testing which items, in sets of two, three or four items, relate collectively to one latent construct. For two constructs on which we have applied our method, we have found that all four and five indicators are most likely common indicators of one construct, respectively. Furthermore, we have found that our method could also be used for discovery or validation of multidimensionality of latent constructs, in a similar style to classical factor analysis.

1.b.iii. We have proposed and evaluated a method for finding the dimensionality of latent constructs within Bayesian network modelling. We have defined dimensionality as a number of states that a latent construct most probably takes, and can also be viewed as cardinality of latent constructs. It can be important for the theory under scrutiny because it might happen that depending on the dimensionality of the construct, the marginal likelihood of the entire model can be different. It can also be useful with respect to the construct itself, because it shows the scale on which the construct operates. For instance, if we conceptualise Satisfaction, we could find out whether it is a dichotomous variable, and takes only two states "low satisfaction", and "high satisfaction", or it spans rather over more intermediate values, e.g., "low satisfaction", "moderate satisfaction" and "high satisfaction". From the modelling perspective, this is also a vital issue, since it can have significant effect on the performance of the model and on its complexity [Elidan and Friedman, 2001]. For two constructs for which we have applied this technique, we have found that both concepts are best represented as ternary variables.

1.c. Another sub-goal that has to be achieved with this work is the examination of Bayesian networks in specific issues in theory development.

1.c.i. First, we examined the ability for modelling moderating effects. Let us recall our discussion from Chapter 5. Assume that a focal variable is directly dependent on two parent variables, one of which is an explanatory variable, and the other one is a potential moderating variable. Consider first the marginal conditional probability distribution of the dependent variable given a specific level of the explanatory variable, and conditionally on one state of the moderating variable. Now, it is legitimate to speak of a moderating effect if there exists a clear difference between this considered distribution and another distribution of the explanatory variable, conditionally on other levels the moderating variable. Of course, the more difference between these distributions, the more significant the moderating effect. Furthermore, we can even study a moderating effect for different levels of the explanatory variable, and observe whether the moderating effect gets stronger or weaker. By means of this reasoning, we have been able to discover a theoretically likely moderating effect of Gender on the link between Ease of navigation and Likelihood to return; nevertheless, to be precise, the actual existence of this effect is, given the data used, less likely than its absence.

In conclusion, the important issue of moderators in the context of CS&L research can be successfully traced and accounted for by the Bayesian network approach. By the way, we note that the analysis of moderating effects with the SEM approach would require two or more different models and would be usually more difficult to reveal [Gefen *et al.*, 2000; Bagozzi and Yi, 1989].

1.c.ii. Furthermore, we have examined Bayesian networks for the ability of discovering and modelling mediating effects. In the Bayesian network framework, deciding whether a variable is a mediating variable or not consists in consulting the marginal loglikelihood of different parents' set for a variable we want to explain, and for a potential intermediary variable. What's more, we can resort to the marginal likelihood of different parents' sets to test for the consequences of omitting intervening variable.

Such an analysis performed in Chapter 4, and plausible theoretical insight was sufficient to conclude that User's Attitude towards the website could be regarded as an example of an intervening variable. We have found that the Attitude can be thought of as the mediating variable as it best explains the entire influence of Ease of Navigation and perception of Look and Feel on the Likelihood to return.

In conclusion, the Bayesian network approach makes it possible to explore the effects of mediating variables and moderators.

Summary

Let us summarise the most important results of the part on the use of Bayesian networks in theoretical e-satisfaction and loyalty research.

We have found that the Bayesian network approach gives theoretically sound results of causal inference from data. In particular, 1) the results of theoretical modelling on the basis of four different datasets in the first case study yielded similar findings suggesting the existence of an overall theoretical model of the e-loyalty; 2) the results of the second case study in customer loyalty corroborate the *a priori* postulated theoretical model of this phenomenon. This suggests that the model validation procedure based on the posterior probability of the model is a valuable way both of discovering and corroborating the theory of Customer Satisfaction and Loyalty. Moreover, both the inductive and deductive approach proved suitable with the use of Bayesian networks.

<u>2. How can marketing theories discovered with BNs be scientifically justified (validated)?</u>

We have investigated the Bayesian network formalism for the fulfilling of the requisites of theoretical models to describe, predict and explain the phenomenon under scrutiny.

In particular, in the context of Bayesian networks, the descriptive power manifests itself both in the qualitative dimension and the quantitative dimension of the model. With respect to the qualitative dimension, description of the e-loyalty phenomenon involves all the names and conceptualisations of the nodes in the model, and the presence or absence of relationships among them, whereas prior marginal probabilities of each of the nodes can be viewed as the quantitative description of e-loyalty. We could consider the chain formula for the joint probability distribution in BNs also as a form of description.

Furthermore, we have verified the potential of the Bayesian network methodology for explanatory modelling. Most importantly, we have found that e-loyalty can be well explained with the perception of Ease of Navigation along with the Attitude. To be more exact, the behavioural dimension of e-loyalty, i.e., Stickiness, can be explained better with the Ease of Navigation, whereas the Intention to return to the website can be best explained both by the Ease of navigation and the Attitude.

We have also verified and confirmed the explanatory power of the models by four criteria: pragmatism, intersubjective certifiability, empirical contents, and by showing that the phenomenon to be explained was expected to occur. We acknowledge that it is difficult to answer what should be the criteria for a satisfying level of the scientific explanation; however, if we accept the weak falsifiability criterion [Hunt, 1991], then our Bayesian network approach can be deemed satisfactory explanation.

Subsequently, we have evaluated the models as predictive systems. Taking into account that customer loyalty is in general a variable, whose state it is difficult to predict, we can say that the predictive accuracy that we have received is quite good. Also, the higher quality of prediction expressed with the Brier score allows us to conclude that the learned models perform reasonably well as classifiers. The classification accuracy is even higher than in other classification-specific techniques.

Summary

In general, we can conclude that the Bayesian network approach can be regarded as a technique that delivers scientifically valid theory. We have found that this methodology is suitable and can be used to model the Customer Satisfaction and Loyalty theory and deliver the scientific understanding of this phenomenon on the empirical basis.

<u>3. What is the added value of modelling marketing problems with Bayesian networks?</u>

3.a. We argue that the added value of Bayesian networks manifests itself with the ability of performing probabilistic reasoning (forward, backward, inter-causal) in the domain.

We have examined the potential of performing probabilistic reasoning (forward, backward, inter-causal) in the e-loyalty domain in Chapter 4. Probabilistic reasoning can be achieved by instantiating constructs to desired states as evidence, and determining the posterior distributions for some variables in focus.

3.b. Next, we investigated and illustrated the potential of performing what-if simulations. An end-user of a Bayesian network-based theoretical model of CS&L can perform what-if simulations. This potential is a consequence of the ability of performing various kinds of probabilistic reasoning in one analysis. Hence, we can find the marginal distribution of any construct conditional on values of its antecedents as well as its consequences in the model at the same time. Another useful type of what-if simulations is entering likelihoods for a variable instead of instantiating it; these likelihoods cause that the variable receives new marginal distributions that we can view as desired prior marginals; now, we can read off the marginals of other variables resulting in this new marginals. In this way, we can, for instance, find out what would be marginals of loyalty on average if customer perceptions were more favourable.

3.c. Another aspect of modelling with Bayesian networks that can be seen as the added value is the potential of combination of prior knowledge with data.

By prior knowledge we mean our beliefs, or theoretical insights, concerning character of specific conditional distributions for each combinations of a focal construct's parents' values. These prior beliefs are then faced with observational data from our study to determine the posterior estimates of the probabilities defining these conditional distributions.

In Chapter 4, we have been able to make use of our knowledge in defining prior ordering of variables. In Chapter 5, we have designed two experiments in which we imposed different priors on parameters of these local distributions. These priors can be seen as "uninformed" in the sense that they do not represent any concrete prior knowledge; the two models examined in these experiments were different from each other in the amount of our ignorance. We observed that they have indeed an effect on the posterior distributions, and even on the marginal likelihood of the model. In our experiment we have found that these priors, even more importantly, have an effect on the relative probability between models. This is probably because there is not much data, and especially because there is no data for the hidden nodes.

In conclusion, we must note that this kind of introducing prior knowledge into the development of theory of phenomenon under study is characteristic of the Bayesian data analysis. This type of analysis can be especially useful when important accumulated knowledge exists with respect to the specific character of the relationship, that we want to account for, between two adjacent constructs, or when data at hand are scarce, or when data come from sources of different kinds.

Summary

In summary, we argue that the unique features of probabilistic reasoning and what-if simulations, as well as the potential of combining existing knowledge about the CS&L phenomenon with data for improved theory discovery and validation, constitute the essential added value of the Bayesian network theoretical models, apart from the necessary capabilities as developing and validation of theory.

<u>II. Conclusions with respect to the use of Bayesian networks in practical Customer Satisfaction studies</u>

In the second part we were concerned with the use of Bayesian network methodology in practical, business-oriented applications. In this part, we have put only one general research question.

<u>1. How can Bayesian networks be applied in service feature/dimension importance/performance study?</u>

1.a. In Chapter 6, our objective was to adapt and examine Bayesian networks in the classification of service dimensions.

1.a.i. First, we demonstrated how Bayesian networks could be applied in service dimensions analysis for identifying the derived importance of service dimensions for overall (dis)satisfaction judgments,

Since satisfaction at the dimension level was not operationalized by the customer questionnaire, variables reflecting service dimensions were created by inferring their values by k-means clustering algorithm based on satisfaction with specific features within the dimension. These variables were binary. We found that the clusters were well separated and could be considered as groups of high and low satisfaction. For these new variables, for each case in the dataset we assigned a value reflecting the level of satisfaction with the service dimension. Next, for both high and low level of overall satisfaction, we expressed their probabilities in terms of the probability of high and low levels of satisfaction with each dimensions, respectively. The procedure we proposed for this purpose is based on the one-way sensitivity analysis in the model, in which the dependent probability is the probability of high, and low, overall satisfaction, and the parameters are marginal probabilities of high and low satisfaction with each feature. We showed that this probability could be illustrated graphically with linear functions.

These graphs confirm the findings in Mittal *et al.*, [1998] in that they show the diverse nature of the influence of satisfaction with a feature on overall service satisfaction: low levels of satisfaction are found hardly sensitive to dissatisfactory experiences with service dimensions, whereas high overall satisfaction shows in this respect an increased dependence.

We found that customer service can be classified as satisfier/dissatisfier. Similarly, we can classify billing also to the same category; nevertheless, billing quality has a more substantial impact on satisfaction than customer service has. Tariffs, due to their positive impact on moderate satisfaction and negative impact on high satisfaction, warrant a closer look to arrive at the right conclusion. We can conclude that the proposed approach is suitable for the analysis off derived importance of service dimensions.

1.a.ii. Furthermore, we have developed a procedure and evaluated Bayesian networks with regard to supporting marketing decisions by means of the importance/performance analysis. Based on the strength of the influence we classified the service dimensions into categories of importance, and augmented with their performance, we carried out an analysis of priorities for improvement.

In order to calculate the performance of the service dimensions, we compared their marginal probability distributions with the one for overall satisfaction, and we found that the performance of all the three dimensions can be classified as low.

From the importance-performance analysis it follows that the company should undertake some actions to improve the performance of the considered service aspects billing being the first priority, and customer service being the second. Further insight regarding phone tariffs is required to formulate a relevant marketing policy in this respect.

We conclude that the presented sensitivity analysis-based approach with Bayesian networks can be used for importance/performance analysis concerning service dimensions.

1.a.iii. We examined also Bayesian networks in terms of discovering interaction effects (synergy and negation) among service dimensions. We have found it likely that some potential determinants of overall satisfaction do not manifest an apparent influence when considered apart from other factors. It can however at the same time happen to be an important factor catalysing the impact of other service dimensions. Synergy effects that can be observed in this situation may be either positive or negative. Therefore, we included a study of interaction effects among the dimensions.

The procedure we proposed is based on the two-way sensitivity analysis in the model, in which the dependent probability is the probability of high, and low, overall satisfaction, and the parameters are marginal probabilities of high and low satisfaction with each feature.

For instance, we have observed a strong positive synergy between satisfaction with customer service and invoicing, and negative effect between invoicing and tariffs. We can conclude that the Bayesian network approach is very useful in determining interaction effects.

1.b. The next sub-goal in the examination of Bayesian networks in applied CS research was to adapt and examine them in classification of service features. We pursued this sub-goal in Chapter 7.

1.b.i. More specifically, we evaluated first the mediated model of overall satisfaction based on the technique of parent divorcing in the analysis of feature importance. In the mediated model that we have evaluated, customer satisfaction at the dimension level mediates the link between service features and overall satisfaction. The results indicate that the investigated approach does not perform successfully. To be precise, we found that such a model does not allow for reliable classification of features because of the small derived effect of features on general satisfaction.

We found also that the classification is not feasible because the relationships between service features and dimensions are too complex for the proposed Bayesian network technique based on sensitivity analysis. In this respect, too many features manifest negative influence on overall satisfaction. For instance, it turns out that for some features, the more probability of high satisfaction with a service feature, the less probability of high satisfaction with the respective service dimension. A possible explanation is that respondents tend to classify the service dimensions and features in different ways.

1.b.ii. Secondly, we pursued to find out whether in the mediated model it is possible to treat satisfaction with the service dimensions as hidden nodes, and thus optimise the design of a customer questionnaire by not asking about satisfaction with service dimensions. Leaving out the questions concerning satisfaction with service dimensions from the questionnaire would be advantageous for customer-oriented companies, since it would simplify the measurement procedure, positively influence the reliability of the included concepts, and reduce the costs of the satisfaction programs.

We have examined models with two different parameterisations. One of them has been fully parameterised on the basis of observed data. In the other model we treated satisfaction with service dimensions as hidden variables, and we used the EM algorithm to estimate the necessary probabilistic distributions. We found that classification of features is different in each case. The reason for that is that the maximum likelihood estimates of the conditional probability tables estimated with the EM algorithm in the model with hidden service dimensions differ considerably from the respective conditional probabilities in the model with observed service dimensions. However, taking into account the predictive accuracy both models perform equally well.

Due to this complex nature, estimation of hidden nodes and labelling by means of Maximum Likelihood can be cumbersome. Aliasing is a crucial problem in this context, since we found it very difficult to retrieve the correct labels for satisfaction at the dimension level.

Interestingly, we have found that it is not necessary to measure customer satisfaction at the dimension level when the main purpose of CS study by means of Bayesian networks is to predict the level of overall satisfaction.

However, when we are concerned specifically with the feature importance-performance analysis, we have found that derivation of probabilities relating to service dimensions only on the basis of data on service dimensions and overall satisfaction has failed. As a result, the learned probabilities do not allow for the use of the examined model in the feature importance-performance analysis.

1.b.iii. Our last objective was to evaluate the noisy-OR model of overall satisfaction in the analysis of feature importance in which features are direct parents of overall satisfaction.

Because of the non-interactive nature of the influence of features on overall satisfaction and its parameterisation, we have expected that modelling with the noisy OR-gate could be expected to be best suitable for detecting "must-be" or "winner" features. However, disjunctive interaction implies also that we consider only those cases in the data for which responses were in some specific configuration. In practice, we have found that many records cannot be used in parameterisation of the noisy OR-model since too many respondents are satisfied with more than one feature at a time. These cases are actually abandoned in the model parameterisation, which has in turn negative consequences for reliability of parameters. It has turned out that the number of cases in this study was too small to yield reliable patterns on the importance of service features.

In general, two conditions must be fulfilled to successfully determine the category of the feature. Firstly and most obviously, the causal strength must be high. Secondly, we can perform classification of features only if coverage is high enough.

We have found that none of the twelve features included in this study when acting on its own has significant effect on high general satisfaction with the service. Furthermore, we have found that in the case of only two features out of twelve considered, these two conditions are satisfactorily fulfilled. More specifically, we found that Top Technology (PQ1) of products is of great importance in contributing to overall dissatisfaction, and Competitive position of the company in the market (Im6) alike.

Therefore, we conclude that the Noisy-OR model of overall satisfaction have turned out useful only in detection of "must-be" features, as we have not found other categories of features than "must-be" in a reliable way.

Summary

We have proposed a technique based on sensitivity analysis in Bayesian networks that could be used in a service feature/dimension importance/performance study. We have found that it can be used in the importance/performance analysis concerning service dimensions. However, it does not enable successful studying of the impact of service features on overall satisfaction in the importance/performance analysis of service features. Furthermore, we conclude that the proposed technique can be used for discovering interaction effects between service features, and concerning the issue of the questionnaire optimisation, we must conclude that our approach does not allow for not asking about satisfaction with service dimensions.

- <u>With respect to the use of Bayesian networks both in theoretical as well as in practical CS studies</u>

In the course of the research presented in this chapter, we have identified several areas in which the Bayesian network approach manifests its strengths and weaknesses with respect to specific technical and statistical and modelling issues, such as data distributional assumptions, missing data handling, etc. As we formulate it in the objectives in Chapter 1, it was not our aim in this thesis to perform true comparison with other techniques; therefore, most of our conclusions should be corroborated in the competitive setting with other methods and statistical tools.

We have found the following strengths of the Bayesian network modelling approach in the context of CS&L research:

- it can handle missing data in a sound way,
- it seems to perform well when data come from a small sample,
- it enables combination of knowledge with data,
- it offers a method of avoiding overfitting,
- it provides simple output statistics, for which no rule-of-thumbs are required,
- the BN approach is user-friendly and easy to interpret,
- it enables modelling one-item operationalization of constructs,
- it enables making predictions,
- it offers the potential of determining the values of latent constructs,
- it enables testing for omitted constructs,
- do not assume any particular distribution.

As regards weaknesses and drawbacks of the Bayesian network approach in general, on basis of this work we found that the following issues can be seen as weaknesses:

- the requirement of collapsing the number of values,
- the requirement of predetermining the directionality of causal influence in the exploratory model learning,
- inability to undergo the strict confirmation (categorical validation),
- furthermore, the approach incorporating latent constructs is subject to weaknesses, including
  - its inability to control for the measurement error,
  - problems with finding optimum, and
  - problems with calculation of effective dimension.

More precisely, with respect to the strengths, we found that Bayesian networks can handle missing data in a sound way. Even if there are lots of missing data, i.e., if up to 50% of all cases on a specific variable are missing, the approach performs well in the sense that it yields similar theoretical model of relationships. Missing values are imputed on the basis of the entire knowledge (theory) encoded by the model.

Next, we must also address the issue of good performance of the BN approach when dealing with rather small data sample sizes. In our study, the data sizes varied from 140 to 409, and for each dataset, we have received similar results in terms of existence of theoretical relationships between some variables or a lack thereof. In our opinion, it is an advantage that regardless of the sample size, which in our study varied from a small dataset to a medium size dataset, we have been able to receive similar theoretically sound results. However, further investigations with larger data samples, and sub-samples are recommended to corroborate this conclusion and to test the sensitivity of the approach to varying the number of cases.

Bayesian networks enable in an easy way the combination of accumulated knowledge and data. In this case study, we have let the prior knowledge of the possible causal ordering of variables be combined with the data. Some authors can see this potential as an unnecessary burden for the researcher; for others, it will be rather seen as an opportunity to make use of the accumulated knowledge. We should note that the issue of combination of prior knowledge with data is an issue of lively debate between proponents of the Bayesian statistics and advocates of traditional statistics. We leave this debate aside, and we state only that the combination of knowledge is one of the characteristics of the Bayesian network approach.

Subsequently, the Bayesian network approach offers a principled method of avoiding overfitting. This means that the marginal likelihood score by its nature strikes a balance between the complexity of the model and the fit to the data.

At the moment, actually the only goodness-of-fit statistic in use is the marginal likelihood; there is no need to calculate any numerous statistics that are hard to interpret. Therefore, no rule-of-thumbs are necessary to interpret the sufficient value of the marginal likelihood; other methods require in that respect that measures exceed some threshold, which is often arbitrary.

We found that Bayesian networks are user-friendly and easy to interpret; elementary knowledge of statistics on the level of the Bayes' rule and basic theorems in probability calculus are sufficient to interpret the consequences of the model. The measure of posterior probability is intuitive and widely known. We argue furthermore that Bayesian networks do not require any background in advanced mathematics or statistics from the researcher to construct a model; other techniques require in these respects much expertise in advanced topics such as matrix algebra, etc. We expect that little effort is necessitated to communicate the results to non-experts and to get them acquainted with this methodology.

We found that one-item operationalization does not pose any problem to theoretical modelling with Bayesian networks, as the indicator is treated as the latent construct itself; it should be treated as an advantage, since other techniques often suffer from under, or over-identification in this respect; and require at least three observed variables per construct [Steenkamp and Baumgartner, 2000].

Bayesian networks manifest not only predictive capabilities, as thanks to the probabilistic reasoning it is possible to predict, or retrodict posterior marginals for any variable in the model, but also these capabilities show good prediction accuracy.

Furthermore, the approach of handling latent constructs that we developed in the second case study (Chapter 5) offers a possibility to determine the value of the latent construct on account of its indicators.

A direct consequence of the work in this case study is that we can check whether introducing new latent constructs does not increase the likelihood of the model. As an example, let us assume that we have constructed a theoretical model of CS&L loyalty for which we obtained a specific value of the posterior (approximated) probability. Now, we could introduce another construct into this model, by positioning it in the model in a place implied by the conceptual meaning of this construct and our new theoretical hypothesis concerning it; the Bayesian network approach enables calculating the posterior (approximated) probability of this new model. Of course, higher probability implies now that our new theory is more likely than the old one, whereas, accordingly, smaller probability will imply that the new theory is less likely. We note that we have not examined such scenarios in this work, but we recognise such a potential of Bayesian networks with latent constructs.

An advantage of Bayesian networks is that they in general do not assume any particular distribution and can accommodate unusual distributions. Probabilistic network models, of which discrete Bayesian networks are a specific kind, do not require any specific probability distribution of the variables, unlike other approaches, including LISREL and PLS models. We have not made any comparison with those techniques, but we expect that the Bayesian network methodology can be a good alternative to these other methods well established in the marketing research.

Let us now discuss the disadvantages in more detail.

A potential weakness of Bayesian networks is that the number of categories that the variables take on should typically be collapsed. The target number of categories depends on the sample size, and should conform to the rule: the smaller the sample size, the fewer categories there should be. The rationale behind this aggregation is to avoid sparse conditional probability tables, as sparse CPT's have a negative effect on computational feasibility of parametric estimation and validation (Bayesian scoring), as well as on the reliability of specific parameters in the CPT's. This requirement should be seen as a weakness, because it can lead to the loss of potentially valuable information, and can obscure the true results.

Furthermore, a drawback of the inductive-exploratory search for the most likely network structures presented in Chapter 4 is that we need to predetermine the direction of the potential causal influence at the beginning of the research design. This can be seen in a sense as a limitation of the reliability of the findings.

We have also observed in this case study that the posterior probability of the models as the goodness-of-fit measure can be viewed as a weakness in the sense that it does not enable categorical confirmation of the model. Typically, in deductive research, the aim of building a theoretical model is to test it empirically to find evidence as to accept or reject this hypothesized model. This can be termed the *strict confirmatory* modelling [Jöreskog and Sörbom, 1993].

Such a procedure is not feasible by taking the Bayesian network approach. To be precise, it is not possible to confirm a theoretical model in the strict sense, as the marginal likelihood measure, until today, cannot be treated with some form of statistical significance test. Nevertheless, it must be noted that a Bayesian network model can be empirically validated in the strict sense using the constraint-based approach that we briefly address in Section 2.6.2 [e.g., Spirtes *et al.*, 2001].

The proposed method of handling latent constructs is subject to weaknesses. The handling of latent constructs and the measurement model is the focus of active research in the BN community at the moment. The methods that we propose in this chapter are attempts to solve this problem. However, a major drawback of our method of measurement modelling is that it is still not able to control for measurement error. Another weakness that we must realize when applying Bayesian networks with measurement models is that we must use approximations of the marginal likelihoods. These approximations require that we estimate conditional probabilities with the EM algorithm, so all consequences of the use of this algorithm must be also taken into account. An important issue that must be mentioned here is the potential problem of under-identification. More precisely, there is no guarantee, with the Bayesian network approach with latent constructs, of finding the global optimum for model parameters (conditional probabilities); we have not done any investigations in this direction, so we stay cautious with making firm statements about this issue. Furthermore, the requirement of multiple restart of the EM algorithm, or slow convergence, can be seen by some authors as another weakness, although in our opinion this disadvantage can be quite well resolved by methods proposed in the Bayesian network literature [Chickering and Heckerman, 1997, p. 195; Thiesson, 1995; Bauer *et al.*, 1997; Fischer and Kersting, 2003].

Also the calculation of the effective dimension for latent construct models should be recognised as a weakness, since this calculation cannot be performed in every model. In particular, the more variables are treated as true latent constructs in the model, the more difficult it is to obtain the effective dimension.

Summary

In summary, we argue that the Bayesian network approach applied in the context of the CS&L research offers more strengths than weaknesses. From among the most important strengths we can mention: the potential of the theoretically sound handling of missing data, the potential of accommodating many different probability distributions of the data, the easiness of its use and interpretation by non-experts. The most important weakness, in our opinion, is the lack of a fully established procedure of structural and measurement modelling; our approach does not allow for controlling the measurement error, and should be treated rather as an initial attempt that addresses this limitation but it does not solve it completely. Furthermore, the requirement of aggregation of values of variables that can lead to potential loss of valuable information can be considered a weakness.

Based on the research accomplished in this thesis, and upon the investigation of the strengths and weaknesses we can conclude when Bayesian networks should be used in particular. We would recommend using the Bayesian network approach especially when:

- the theoretical constructs can be measured reliably with one item,
- we do not have or do not want to advance any dominating or alternative hypothetical models and would like to perform an inductive, exploratory search in the space of many diverse models for the most likely theoretical structure,
- some data are missing at random or structurally missing,
- we require the model to have theoretically sound assumptions related to the CS&L theory and high practical value,
- we intend to use the model for classification and prediction of different variables,
- we intend to use the model for diagnostic or predictive simulations and what-if analysis,
- the data have multinomial distribution,
- we aim at discovering and validating moderating and interaction effects,
- non-linear effects between variables are very important to capture and to model.

8.2. Implications

From the research accomplished in this work we can draw implications both for researchers engaged in theoretical research on Customer Satisfaction and Loyalty, as well as for practitioners involved in applied Customer Satisfaction modeling. We will review these implications on a chapter-by-chapter basis.

8.2.1. For CS&L researchers

We suggest that Bayesian network can be more widely used in theoretical research on Customer Satisfaction and Loyalty.

First of all, we postulate that the Bayesian network approach in the CS&L research is suitable and performs very well both in the context of discovery of the theory of this phenomenon as well as in the context of scientific justification of the theoretical insights into the domain.

In the context of theory discovery, on the basis of the analysis of visitors of four main portal websites in the Netherlands performed in Chapter 4, the use of Bayesian networks proved suitable in the inductive approach. A negative side of the presented inductive approach is that we predetermined the direction of the potential causal influence at the beginning of the research design. This can be seen in a sense as a limitation of the reliability of the findings. Of course, we can re-validate the results by allowing for other models starting with different search orders. Then, from among all the resulting models, the best model can be chosen on the basis of its posterior probability. On the other hand, a positive aspect of our methodology is that the procedure delivered the most appealing result in the sense that the variables, for which most nodes were tested as potential parents, eventually occurred to be child nodes of the variables located closer in the initial search ordering. This concerns the loyalty variables, as they were following the attitude. Such a result partially confirms the ordering that we have assumed. More importantly, we found that the individual models discovered for four data sets as well as the resulting overall model of e-loyalty are theoretically sound. We can affirm that the presented inductive approach can be quite successfully applied with the Bayesian network methodology.

Similarly, on the basis of the results in Chapter 5, in which we examined our approach in the deductive CS&L research, we postulate that the Bayesian network approach makes theoretically sound inference from data. In particular, the results of the customer loyalty study in this chapter corroborate the *a priori* postulated theoretical model of this phenomenon. This suggests that the model validation procedure based on the posterior probability of the model is a valuable way both of discovering and corroborating the theory of Customer Satisfaction and Loyalty. Moreover, our implementation of the presented deductive approach proved suitable also with the use of Bayesian networks with latent constructs.

We propose that one can use Bayesian networks in order to find the best fitting model using an automated search procedure and to discriminate between models using Bayesian scores, such as marginal likelihood of the model. In this context, an important question that a marketing researcher would often like to ask is how big the difference is in the goodness of fit between alternative theoretical models. To our knowledge, Bayesian network modelling has no other instrumentation to judge over statistical significance other than subjective opinion. When the difference between two models seems insignificant and the predominant aim of modelling is prediction we advice to use Bayesian averaging instead of Bayesian model selection procedure.

The marginal likelihood measure avoids overfitting. We can see that the measure by its nature strikes a balance between the complexity of the model and the fit to the data. By consulting the tables reporting marginal loglikelihoods of different sets of parents, we can become convinced that the marginal likelihood makes a "fair" judgment between configurations of one, two, and three parents, namely by selecting this configuration that is the most probable.

With regard to the context of justification, we have found out based on the study in Chapter 4 that Bayesian network modelling can be successfully applied both for explanatory and predictive research. This is one of the most constructive results. In fact, we argue that the explanatory power of Bayesian networks with respect to CS&L is by far more substantial than of other alternative techniques, such as SEM models. Also the potential of forward, backward and inter-causal prediction, as well as what-if simulations is unique and contributes to much better understanding of the CS&L phenomenon than other techniques.

In Chapter 5 of this work, we have also proposed and examined a method of incorporating the measurement model into causal modelling with Bayesian networks by introducing latent variables operationalized with multi-item measurement scales directly in the model. In particular, we encourage CS&L researchers to apply the proposed approach in their research practice, as our experience delivers very positive results on our approach. Furthermore, we suggest to get familiarised with the method since it enables performing construct validation and finding the best dimensionality of latent constructs. The procedure of construct validation taken in this study aims to assess whether the indicator variables relate to one potential construct, or to more constructs. In our method, we consider dimensionality as the most likely number of values that a latent construct takes on. Moreover, we have found that aliasing does not pose any problem, since the meaning of states of latent constructs can easily be established from the indicators. We are also quite convinced that even when all concepts are measured with one-item measures, as it was the case in Chapter 4, the Bayesian network approach is useful. All in all, the results in all these issues are very constructive.

Simultaneously with the procedure of construct validation, we can check and discover whether introducing new latent variables does not suggest existence of new, or omitted, latent constructs. In that case, if the network structure augmented by the introduction of a latent construct (of course without their equivalent indicator variables) would represent higher value of the likelihood, then this might be an indication that this new, previously not considered construct, can potentially play an important role in the theory under consideration. By looking at relationships between this new construct and the remaining constructs, we can also get an idea what omitted concept the construct should represent [see e.g., Heckerman *et al.*, 1999]. This capability presents itself as an advantage over SEM modelling.

Given the positive results of the search in the space of hypothetical models, examined in Chapter 4, and additionally taking into account the potential to locate "unknown" omitted latent constructs described above, we propose that the Bayesian network approach can be useful also for a special kind of exploratory CS&L research.

We have presented two scenarios that a theoretical CS&L researcher can be faced with: the model-generating scenario in Chapter 4, that can be viewed also as an inductive or exploratory scenario, and the alternative, or competing, models scenario in Chapter 5; we conclude that both scenarios can be successfully achieved with the Bayesian network methodology. However, often, a researcher is only interested whether a particular hypothetical model should be accepted or rejected on the basis of the data. We acknowledge that the potential of the strict confirmation of a model in such a situation, also known as the dominating hypothesis scenario, can be for obvious reasons advantageous to the researcher, but it can be also easily criticized, because often several different models are supported by the data. Anyway, the Bayesian score approach to theory discovery with Bayesian networks, taken in this research, fails in the strictly confirmatory scenario. The reason for that failure is that the Bayesian score approach applied here operates inherently with the notion of model's probability, and such an approach is not subject to any categorical, clear-cut statements of empirical adequacy of a model. However, in the strictly confirmatory research scenarios, we argue that this limitation can be mitigated by taking the constraint-based method of validation [see Section 2.6.2. for details on this method].

Both in Chapter 4 and Chapter 5, we have demonstrated that missing data poses no problem for the proposed methodology when estimating the parameters of the model. By means of the EM algorithm, missing values in the network can be imputed in a very sound way by using all the knowledge, or theory, that the model represents. For example, even when a particular respondent has responded to one question in a survey, it is very easy to make use of this single datum, and to estimate the most likely values of other variables for this respondent (by means of reasoning in the model); naturally, added value of this particular case in the model estimation is typically negligible, but by this example we would like to point that missing data poses no problem. This is an interesting implication for researchers faced with bad quality data since often they are forced to leave out the cases with missing values, which can contribute to less powerful tests of significance and impair the quality of their work. Furthermore, even when the model is ready to use, it is perfectly feasible to adapt this existing model in the light of new data.

Lastly, besides handling missing data, we argue the Bayesian network based modelling offers many other advantages for theoretical CS&L researchers and overcomes some of the deficiencies of other similar modelling approaches. These strengths are presented in more detail in the previous section on conclusions. Of interest to the CS&L researcher is also the issue when to use Bayesian networks in particular that we also addressed in the previous section.

8.2.2. Implications for managerial practice

Our experience gained with research presented in this work shows that modelling with Bayesian networks offers high practical value.

First of all, we believe that the analytical capabilities of the Bayesian network approach, including the capabilities of what-if simulations and forward, backward and inter-causal probabilistic inference can prove useful for marketing managers in practice. These capabilities provide managers with the technique to predict future behaviour and to ask diagnostic "what if" questions based on assumed marketing actions.

What's even more important, even when applied marketing modellers do not possess enough theoretical insight in order to design a model for a specific problem, they can make use of the search algorithm that will determine the most likely model. In this case, they will need a set with observational and response data, of course. Positive point in this context is that they can use one-item scales and do not need to worry too much about low response rates, or missing data, whenever they need more theoretical insight, or plan conducting customer satisfaction programmes.

All these capabilities come along with the potential of developing and validation of the theory of the marketing phenomenon in focus.

All in all, this means that managers are offered a powerful technique that, on one hand, allows for introducing theoretical insight into the model and has high explanatory value, and on the other hand, a technique that has high pragmatic value for managerial practice.

Furthermore, the findings presented in Chapter 4 provide insight into the theory of e-loyalty that has high practical value not only to applied marketing modellers but also to web marketing managers. For example, one of the most surprising results that we have found rather unexpectedly is that the general attitude towards a portal site is not as important as the perception of ease of navigation in the formation of customer e-loyalty. Hence, we would like to stress the importance of easiness of navigation, especially while designing portal sites.

We have found that the joint probability distribution of the variables in the customer e-loyalty phenomenon can be best represented with a probabilistic dependency structure in which visitor's sociodemographic profile is not relevant with any other variable. The findings suggest that age and gender are determinants of position in the household, which is, on theoretical grounds, a plausible result. We argue that it does not make sense to segment visitors according to these attributes in other customer e-loyalty studies.

Furthermore, we have found that, unsurprisingly, visitor opinions matter to a great extent. From the three opinions on website characteristics that we considered, visitor opinion about the ease of navigation seems to be the most important one.

From Chapter 5 some recommendations for marketing managers concerning customer involvement and loyalty can also be drawn. For instance, from our finding that given high Trust there is more probability of high Loyalty than given high Involvement (this effect is stronger), we can recommend that the companies should stimulate high confidence of their customers rather than their engagement.

Next, we argue that practitioners will find the presented approach valuable, as unlike it is the case with other techniques, it easily enables to determine the value of the latent construct based on the values of the indicator variables. As a result, they can perform simulations by assuming some values of the observed variables, introducing this information as evidence into the model, and they can find out the posterior distribution for the corresponding latent variables by performing reasoning in the network. Even more interestingly, they can see the effect of these assumed values of indicators on other constructs in the network. We believe that this capability is of great value to marketing managers.

Another important implication of the research in Chapter 5 for marketing managers can be that they will find the use of latent construct Bayesian network models easy and intuitive. They should find it easy to advance several competing structural models, link the latent constructs to their indicators, and draw conclusions from comparison between these models. This finding should yet be corroborated in practice by exposing our approach to managers and marketing practitioners.

Summarizing the use of Bayesian networks in a study of dimension importance/performance examined in Chapter 6, we can recommend that managers and applied CS marketing modellers can apply the methodology to classify service dimensions. In this study we concluded that customer service could be classified as satisfier/dissatisfier. Similarly, we can classify billing also to the same category; nevertheless, billing quality has a more substantial impact on satisfaction than customer service has.

From a managerial perspective, outcomes of the present technique seem to be of interest, as they indicate which dimensions should be taken care of, and which of them are less important and deserve less attention. For instance we found, that the company should undertake some actions to improve the performance of the considered service aspects billing being the first priority, and customer service being the second. Further insight regarding phone tariffs is required to formulate a relevant marketing policy in this respect. Classification of service features is however more difficult.

It is also possible to find out the synergy and negation effects, if exist, between perception of different service dimensions. We can conclude also that the Bayesian network approach is very useful in determining interaction effects.

We think that managers should familiarize with the Bayesian network modeling approach. The advantage of Bayesian networks is that they are easy, intuitive in use and do not require any expertise in understanding the results. The results have a probabilistic nature and, unlike other causal modelling techniques, are easy to interpret. All the relationships are viewed probabilistically, thus allowing for easy interpretation. The outputs of this analysis are of a probabilistic nature and easy to interpret for managers.

We posit that more applications aimed to support decision-making in companies will be built by use of the Bayesian network technology in data analysis and consultancy.

On the basis of Chapter 7, in which addressed the importance/performance analysis of features, we suggest that indirect derivation of satisfaction with service dimensions entirely on the basis of response data on overall satisfaction score, and satisfaction with features turns out not to be successful. The nature of the relationships between performance of service features and the judgments of overall satisfaction is probably too complex for the ML optimisation to approximate the original relationships between features and overall satisfaction.

The finding that the categories of two models, one with observed dimensions and the other one with dimensions estimated, are different implies that we should operationalise also satisfaction at the dimension level in future customer satisfaction studies.

However, on the other hand, the results suggest that asking directly for some compound construct is not necessary if the main model's use is the prediction of overall satisfaction, since the same predictive power of the model can be obtained when the satisfaction with dimensions is treated as hidden. Otherwise, if the model is intended to be used in feature importance/performance study, it must be used with a lot of caution.

8.3. Limitations

The overall research objective accomplished in this work, that is, the critical examination of the application of the Bayesian network approach in the CS&L research and the development of new methods within this approach to improve its current abilities, is subject to limitations that we will now address. These limitations will be discussed for each case study at a time.

One of the main limitations of the case study in Chapter 5, in which we examined Bayesian networks in inductive research on the example of the e-loyalty phenomenon, is a requirement of the prior ordering of variables. The results of a study by Chickering *et al.* [1995] suggest that the greedy algorithm that we applied is sensitive to variable ordering, so that the specification of the prior ordering can influence the results to a large extent. Of course, we can re-validate the results by allowing for other models starting with different search orders. Then, from among all the resulting models, the best model can be chosen on the basis of its posterior probability. We haven't performed experiments with another initial orders of variables, which clearly is a limitation of the examination of the Bayesian network approach. There are various approaches to circumvent this limitation. For instance, we could use the more time-costly edge-reversal search procedure that does not require an ordering. Other efforts are directed at the selection of the initial ordering, for instance, Larrañaga *et al.* [1996] use genetic algorithms to obtain the best ordering of the variables. This issue can be a topic for further research.

From the perspective of the e-loyalty theory that we considered in Chapter 4, we agree that the concept of e-loyalty operationalized by stickiness and intention to return can have some drawbacks. Namely, the behavioural aspect might not be well accounted for by our conceptualisation. Stickiness might not be an objective measure of behavioural e-loyalty, since according to our operational definition it implies that a user that has visited the site only once for a long time, is more loyal than a user that visits regularly but shorter on average. Furthermore, it might be dependent on the Internet connection speed (bandwidth) and other factors; therefore the model we developed has a limited theoretical significance, as many important concepts are left out.

The predictive power of the models in the case study in Chapter 4 was tested only for one particular variable, i.e., Attitude. We acknowledge that the capability of the theoretical models to predict should be ideally tested for more variables in order to obtain more reliable judgment in this respect. Nevertheless, the results that we present here for predictions of only one variable seem reasonably promising.

A potential threat to validity of our results, and thus a limitation of the results of the examination of the Bayesian network approach, especially for the fact that all four data sets in this case study yield very similar theoretical relationships is that the data sets have many missing values. For example, the dataset that describes users of the Ilse portal reports as much as 49.3% of missing data on Ease of navigation. This could potentially have a negative effect on the value of the used Bayesian score and missing data handling of Ramoni and Sebastiani [1997] in the sense that variables with many missing values could be given more likelihood as parents. Although at the first sight, this effect is quite likely given our results and should be taken into consideration, we haven't found any convincing evidence that this effect is significant; moreover, the method is believed to be robust with respect to missing values [Ramoni and Sebastiani, 1997].

So as to leave no doubt, it must be noted that any Bayesian network model that is validated on data should be viewed as explanatory for the theory under consideration to the extent that it explains these data, and not that it explains the process or the phenomenon. Of course, the more the model is rooted and supported by the existing body of research in the discipline in question, the more confidence we can have that the model also concerns indeed the "true" theory.

Another limitation of the presented methodology is its inability to undergo the categorical validation, i.e., a Bayesian network model cannot be validated, unless it is compared with alternatives. It is so because we get a posterior probability over models we consider. That means that we cannot accept the learned model in isolation from other models. We could accept the learned model if its probability is significantly higher than any other alternative model, as is the case in Bayesian model selection. In case the best model is not remarkably better than others we should not be overconfident in the model. The problem that arises is therefore how to judge if the difference between models is big enough. This decision is usually taken on a subjective basis and should be addressed in future work.

Research in Chapter 5, in which we introduced the structural and measurement modelling, has its limitations too. Measurement modelling has been originally developed as an instrument of accounting for the measurement error, which should be the explicit component of marketing models [Steenkamp and

Baumgartner, 2000]. In the classical true-score theory of measurement [Lord and Novick, 1968], the observed score equals the true unobserved value plus the error term. From the point of view of this theory, the measurement modelling approach that we presented in this case study can be criticised for departure from this principle of full incorporation of the measurement error in the holistic analysis. A limitation of the proposed approach to measurement modelling could thus be that the measurement error in the relationships between latent construct and the corresponding observed variable cannot be separated qualitatively from the true score for the latent variable. In our approach, this error manifests itself rather in the conditional distribution for the observed variable given the true score on the latent one, and more precisely in the uncertainty around the corresponding state of the indicator. If the measure is theoretically and observationally meaningful, than the uncertainty can be entirely viewed as the error measurement; the more uncertainty, i.e., the more probability mass is distributed to other states of the indicator, the greater the measurement error.

We can conclude that the handling of the structural model as consisting of hidden nodes estimated by the EM algorithm leads to theoretically sound conditional probabilities. However, we do not know exactly what the precise impact is of the EM estimation on the conditional probabilities. We conjecture that conditionals are likely to be too soft and the EM estimation makes the distribution be smoothed compared to "true" conditionals in reality, so they should be taken with caution. This issue could also be an interesting topic for future research.

While discussing construct validation method, we considered existence of only two latent constructs that the indicators could relate to. Therefore, settings in which three and more constructs are present should be tested as well.

A potential serious bottleneck of modelling hidden variable Bayesian networks is the calculation of the effective dimension, which is required to approximate the marginal likelihood of the model. Taking the structural dimension on the other hand can bias the results. The problem of calculation of the effective dimension grows with the number of hidden variables. The more hidden variables there are in the model, the more time it takes to estimate it. Furthermore, another limitation of the approach with latent constructs is that no precise measures of marginal likelihood of the model exist, so that one has to fall back on approximations, such as the Bayesian Information Criterion and the Cheeseman-Stutz score, which might not be precise enough. The performance of these approximations in other service settings and data can be thus addressed in future.

We did our best to ensure that the case studies on the theoretical CS&L research and their theoretical results concerning the scientific insight into the CS&L domain could be seen as reliable as possible. Still, these results should rather be seen as tentative and illustrative of the Bayesian network approach, and inferior to the objectives of this thesis. We acknowledge that only by the subsequent accumulation of findings from other CS&L studies, including the ones presented here, can we attempt to find generalized "laws of marketing", if they exist. We stress also that further application of the BN methodology in other customer loyalty settings and data is recommended to corroborate the added value of this approach. On the other hand, we acknowledge the fact that in the short term the proposed methodology can be useful for managers who want to gain more understanding of their customers and their own business.

Furthermore, the examination of Bayesian network in the part on applied CS research is subject to the following limitations.

A well-known problem that occurs in traditional customer satisfaction studies is that if a list of features included in the investigation becomes too long, then it makes the analysis complicated and unreliable. The models require in this situation too many parameters that cannot be reliably estimated with available data. Alike, one of the limitations of the presented approach in the analysis of service dimensions in Chapter 6 is that it is also not feasible to study the interaction of many dimensions at the same time because the conditional probability table is growing very quick with the number of features, and causes nuisance with the model's parametric estimation. Furthermore, satisfaction with service dimensions was in this case study created artificially by finding two clusters of users in terms of their satisfaction with features relevant to each dimension. We should test how the Bayesian network technique will perform in if satisfaction with dimensions is also operationalized by the questionnaire and included in the model as observed variable.

The main limitation of research presented in Chapter 7, i.e., the examination of the mediated and the noisy-OR model of overall customer satisfaction, is that these methods still allow studying a very limited number of service features at a time in a feature performance/importance study. Therefore, in order to be able to carry out the analysis we selected a small number of features randomly. We should take into account that the three features that we selected might not be best determinants of satisfaction with their respective service dimension.

In both case studies on practical customer satisfaction research, i.e., in Chapter 6 and 7, we assumed that satisfaction judgments with respect to different service features are marginally independent with each other. Similarly, we assumed also the marginal independence between satisfaction judgments with different service dimensions. In practice, judgments both between satisfaction judgments levels can be interdependent and related with each other, so we conjecture that this assumption could have influence on the results of the importance/performance study performed in these chapters, so further analysis with models refraining these restricted assumptions.

8.4. Directions for further research

Let us first sketch a few avenues for future research in terms of Bayesian network modeling in theoretical CS&L studies.

First of all, a thorough comparison of Bayesian networks in a competitive setting with other techniques, especially Structural Equation Models and Partial Least Squares is of most interest. This comparison could be for instance achieved in a Monte Carlo simulation study, in which we used a data set generated from a causal integral model of customer loyalty. Such a model could be constructed taking into account existing theoretical insight of a panel of experts in the domain. One of the criteria to compare these approaches could be the accuracy with which each model recovers the original model proposed by the experts. Of course, different settings can be considered, such as varying amount of data, varying the frequency of missing values, various strengths of theoretical relationships, etc.

One of the aspects of the work in Chapter 4 that should be addressed in future work is finding a method that makes the specification of the prior ordering of variables unnecessary. Some potential methods in this respect include genetic algorithm-based search for the best ordering [Larrañaga *et al.*, 1996; Hsu *et al.*, 2002].

Another topic for further work is to analyse the impact of different schemes of collapsing the number of categories of observed variables to a manageable number on the results of structural learning, in terms of favouring the existence of links between constructs or the lack thereof. Similarly, studies of its impact on the strength and the character of these relationships, and on the reliability of parameters should also be undertaken. Especially, the issues of applying the equal frequency binning principle and of the optimal reduction scheme are of significant importance in this respect.

As regards the case study in Chapter 5, we stress that further application of the Bayesian network approach with latent constructs in other customer satisfaction and loyalty settings involving diverse data sets is recommended. The recognition of Bayesian networks as a fully legitimate techniques for theoretical modelling requires that issues like reliability and validity are fully taken account of and attainable within the scope of the technique. Of these topics, in this chapter we have presented a possible measure for assessing construct validity, but this and other topics in these respects call for more attention.

One of the most important suggestions for future examination is analysis of the behaviour of the EM estimation in terms of the conditional probabilities and marginals. It would be very interesting to carry out studies on simulated data.

Another interesting topic for further exploration is the analysis of statistical characteristic and behaviour of the presented method of construct validation. For example, in our procedure of construct validation in Section 5.5.1, we performed validation of each measurement instrument in isolation from the complete model. However, the validation of the instruments could also be achieved by considering them in the broader contexts of the entire model, as it could turn out that the mutual relationships in the model play a role in assessing the impact of latent constructs on the indicators. Since we have tested the proposed approach of construct validation only on two latent constructs, it is too few to give any solid assessment. Hence, this method should be merely seen as an initial attempt directed at developing a construct validation procedure within the Bayesian network framework. Therefore, further thorough investigation of properties of our method is necessary in follow-up studies. Various measurement instruments already validated by other authors and well established in the literature should be used as test instances. Further evaluation of this method could be based on comparison with the standard methods applied in SEM modelling, such as multitrait-multimethod (MTMM) of Campbell and Fiske [1959].

Further work is required to corroborate the correctness of the presented approach of finding dimensionality of latent constructs. Central issue is whether models that postulate three states of latent constructs could be preferred over models having other number of states than three simply by the fact that the indicators are also ternary. So, further enquiries are warranted in this respect, for instance by observing the effect of variation of the cardinality of the observed variables from two to the original value of ten.

Thanks to recent advances in structural learning of Bayesian network models from data, methods have been proposed that facilitate finding most likely models with latent constructs directly from the data by means of efficient search algorithms [e.g., Russel *et al.*, 1995; Friedman, 1998]. The common motivation for these methods is that bringing in a new variable can simplify and compact the structure of the model. As the central feature of these methods, during the search for the most likely model, it is evaluated whether there could be any potential hidden variables in the domain, i.e., variables that are not present in the observed data. Roughly speaking, this is done by hypothesising the presence of a latent variable at a certain place in the model, and if the marginal likelihood of such an augmented structure is higher than the one of the original structure, then this variable is retained in the model. Its theoretical meaning can be then guessed on the basis of the location and relationships with other constructs. Further enhancements of these approaches and corroboration of their use in the exploratory CS&L research is one of very exciting avenues for further scientific work. Other analogous topic could be how the presence of hidden constructs can be detected without the need of scoring the entire model.

Research should also be undertaken to work out methods or ways of calculation of the effective dimension of models with latent variables.

With respect to the use of Bayesian networks in practical CS studies, future research may be focused on investigation of models involving more service dimensions and testing sensitivity of the approach in this respect.

We found that aliasing poses a serious problem to successful classification of features in the mediated model. We speculate that this problem can be tackled by assigning non-uniform prior information to the hidden constructs, and doing the Maximum A Posteriori optimisation rather than the Maximum Likelihood optimisation. By assigning non-uniform prior distribution, we impose states on the values of hidden variables. Furthermore, we speculate that if the CPT tables were smaller, allowing for better use of observed data, then Maximum Likelihood estimation could be more successful. The examination of this assumption can be addressed in future work.

In addition, with respect to the noisy OR-model, a severe limitation of modelling overall satisfaction with this model is that this model assumes non-interaction effects among service features. Therefore, other functional dependencies enabling interaction effects should be also considered, including generalizations of the noisy OR-model, such as noisy-MAX, or noisy-MIN models [Srinivas, 1993; Takikawa and D'Ambrosio, 1999].

Another issue worth investigation is to analyse the sensitivity of classification both with the mediated model as well as with the noisy-OR gate to different aggregation thresholds used to recode the original data.

Last but not least, application and evaluation of Bayesian networks methodology in decision modelling by taking into account also costs of marketing actions could be considered in future. We speculate that this kind of decision support modelling can be achieved with so called influence diagrams, i.e., Bayesian networks augmented with action and cost nodes.

Appendix A

The marginal loglikelihoods (MLL) and the Bayes factors between the dependency of and other dependencies explored during the search in the space of theoretical hypotheses in Chapter 4.

Age

Rank	Potential parents for Age	MLL	Bayes factor
1.		-303.337	1
2.	Gender	-312.652	1.11E+04

Table A.1 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Age and other dependencies explored for Freeler data.

Rank	Potential parents for Age	MLL	Bayes factor
1.		-147.972	1
2.	Gender	-155.364	1.62E+03

Table A.2 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Age and other dependencies explored for Ilse data.

Rank	Potential parents for Age	MLL	Bayes factor
1.		-427.021	1
2.	Gender	-436.556	1.38E+04

Table A.3 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Age and other dependencies explored for MSN data.

Rank	Potential parents for Age	MLL	Bayes factor
1.		-185.158	1
2.	Gender	-190.336	1.77E+02

Table A.4 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Age and other dependencies explored for WOL data.

Education

Rank	Potential parents for Education	MLL	Bayes factor
1.		-284.423	1
2.	Gender	-300.790	1.28E+07
3.	Age	-327.262	4.03E+18

Table A.5 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Education and other dependencies explored for Freeler data.

Rank	Potential parents for Education	MLL	Bayes factor
1.		-263.184	1
2.	Gender	-282.169	1.76E+08
3.	Age	-295.376	9.57E+13

Table A.6 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Education and other dependencies explored for Ilse data.

Rank	Potential parents for Education	MLL	Bayes factor
1.		-717.402	1
2.	Gender	-736.164	1.41E+08
3.	Age	-741.121	2.00E+10

Table A.7 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Education and other dependencies explored for MSN data.

Rank	Potential parents for Education	MLL	Bayes factor
1.		-246.173	1
2.	Gender	-261.202	3.37E+06
3.	Age	-279.974	4.78E+14

Table A.8 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Education and other dependencies explored for WOL data.

Position in the household

Rank	Potential parents for Pos_Household		MLL	Bayes factor
1.	Education		-204.714	1
2.	Age		-212.338	2.05E+03
3.	Education	Gender	-216.390	1.18E+05
4.	Education	Age	-218.304	7.98E+05
5.	Gender		-236.399	5.76E+13
6.			-245.770	6.76E+17

Table A.9 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Pos_Household and other dependencies explored for Freeler data.

Rank	Potential parents for Pos_Household		MLL	Bayes factor
1.	Age		-136.493	1.00E+00
2.	Gender		-139.392	1.82E+01
3.			-142.927	6.23E+02
4.	Age	Gender	-143.135	7.67E+02
5.	Education		-176.673	2.82E+17
6.	Age	Education	-198.583	9.23E+26

Table A.10 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Pos_Household and other dependencies explored for Ilse data.

Rank	Potential parents for Pos_Household			MLL	Bayes factor
1.	Age	Gender		-373.379	1
2.	Age			-378.529	1.72E+02
3.	Gender			-384.831	9.41E+04
4.				-408.757	2.32E+15
5.	Education			-438.386	1.71E+28
6.	Age	Education		-484.782	2.41E+48
7.	Gender	Age	Education	-526.639	3.63E+66

Table A.11 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Pos_Household and other dependencies explored for MSN data.

Rank	Potential parents for Pos_Household	MLL	Bayes factor
1.	Gender	-176.567	1

2.	Gender	Age	-177.565	2.71E+00
3.	Education		-187.425	5.20E+04
4.	Age		-190.098	7.52E+05
5.	Gender	Education	-190.807	1.53E+06
6.			-194.590	6.72E+07

Table A.12 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Pos_Household and other dependencies explored for WOL data.

Look & Feel

Rank	Potential parents for Look_Feel		MLL	Bayes factor
1.	Layout		-161.893	1
2.	Layout	Gender	-175.300	6.65E+05
3.	Education		-177.624	6.79E+06
4.	Layout	Education	-182.151	6.28E+08
5.	Layout	Pos_Household	-188.313	2.98E+11
6.			-189.738	1.24E+12
7.	Gender		-195.909	5.93E+14
8.	Pos_Household		-201.178	1.15E+17
9.	Age		-203.142	8.22E+17
10.	Layout	Age	-204.036	2.01E+18

Table A.13 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Look_Feel and other dependencies explored for Freeler data.

Rank	Potential parents for Look_Feel		MLL	Bayes factor
1.	Layout		-129.568	1
2.	Pos_Household		-132.139	1.31E+01
3.			-138.421	7.00E+03
4.	Layout	Pos_Household	-141.672	1.81E+05
5.	Layout	Gender	-143.858	1.61E+06
6.	Gender		-143.869	1.63E+06
7.	Age		-150.151	8.70E+08
8.	Layout	Age	-156.363	4.34E+11
9.	Education		-166.953	1.72E+16
10.	Layout	Education	-202.929	7.26E+31

Table A.14 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Look_Feel and other dependencies explored for Ilse data.

Rank	Potential parents for Look_Feel		MLL	Bayes factor
1.	Layout		-299.296	1
2.	Layout	Gender	-320.579	1.75E+09
3.	Layout	Pos_Household	-327.288	1.43E+12
4.	Layout	Age	-345.033	7.29E+19
5.	Pos_Household		-369.429	2.87E+30
6.	Layout	Education	-386.454	7.11E+37
7.			-400.220	6.77E+43
8.	Gender		-406.906	5.42E+46
9.	Age		-415.955	4.62E+50

10.	Education	-423.604	9.69E+53

Table A.15 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Look_Feel and other dependencies explored for MSN data.

Rank	Potential parents for Look_Feel		MLL	Bayes factor
1.	Layout		-122.034	1
2.	Layout	Gender	-129.123	1.20E+03
3.	Layout	Education	-137.516	5.29E+06
4.			-141.409	2.60E+08
5.	Education		-144.146	4.01E+09
6.	Gender		-146.138	2.94E+10
7.	Layout	Pos_Household	-146.537	4.38E+10
8.	Layout	Age	-151.044	3.97E+12
9.	Pos_Household		-152.397	1.54E+13
10.	Age		-154.002	7.65E+13

Table A.16 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Look_Feel and other dependencies explored for WOL data.

Ease of navigation

Rank	Potential parents for Navigation		MLL	Bayes factor
1.	Look_Feel		-156.033	1
2.	Layout		-160.190	6.39E+01
3.	Education		-160.559	9.23E+01
4.			-168.646	3.01E+05
5.	Look_Feel	Gender	-169.302	5.79E+05
6.	Gender		-174.647	1.21E+08
7.	Look_Feel	Layout	-176.845	1.09E+09
8.	Pos_Household		-178.365	4.99E+09
9.	Look_Feel	Pos_Household	-182.439	2.94E+11
10.	Age		-186.818	2.34E+13
11.	Look_Feel	Education	-197.301	8.36E+17
12.	Look_Feel	Age	-208.836	8.55E+22

Table A.17 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Navigation and other dependencies explored for Freeler data.

Rank	Potential parents for Navigation	MLL	Bayes factor
1.		-79.349	1
2.	Layout	-82.456	2.23E+01
3.	Gender	-83.434	5.94E+01
4.	Look_Feel	-84.308	1.42E+02
5.	Pos_Household	-88.010	5.77E+03
6.	Age	-90.325	5.85E+04
7.	Education	-95.812	1.41E+07

Table A.18 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Navigation and other dependencies explored for Ilse data.

Rank	Potential parents for Navigation	MLL	Bayes factor

1.	Layout		-315.148	1
2.	Look_Feel		-322.152	1.10E+03
3.	Pos_Household		-328.857	9.00E+05
4.	Layout	Gender	-333.961	1.48E+08
5.	Layout	Look_Feel	-336.607	2.09E+09
6.	Layout	Pos_Household	-342.373	6.67E+11
7.			-345.111	1.03E+13
8.	Gender		-348.148	2.15E+14
9.	Age		-359.233	1.40E+19
10.	Layout	Age	-362.044	2.33E+20
11.	Education		-372.852	1.15E+25
12.	Layout	Education	-413.234	3.97E+42

Table A.19 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Navigation and other dependencies explored for MSN data.

Rank	Potential parents for Navigation		MLL	Bayes factor
1.	Look_Feel		-134.154	1
2.	Layout		-137.297	2.32E+01
3.	Look_Feel	Gender	-148.732	2.15E+06
4.	Education		-150.498	1.25E+07
5.			-153.740	3.21E+08
6.	Look_Feel	Education	-155.982	3.02E+09
7.	Look_Feel	Layout	-156.156	3.59E+09
8.	Gender		-158.887	5.52E+10
9.	Look_Feel	Pos_Household	-160.869	4.00E+11
10.	Pos_Household		-162.924	3.12E+12
11.	Look_Feel	Age	-165.631	4.68E+13
12.	Age		-167.321	2.54E+14

Table A.20 The marginal loglikelihood (MLL) and the Bayes factor between the dependency of Navigation and other dependencies explored for WOL data.

Appendix B

Conditional probabilities in models in Chapter 4.

Gender

	Freeler	Ilse	MSN	WOL
Counts	215	140	409	169
male	0.707	0.663	0.713	0.691
female	0.293	0.337	0.287	0.309

Table B.1 Prior marginal probabilities for Gender.

Age

	Freeler	Ilse	MSN	WOL
Counts	215	140	409	169
< 19	0.302	0.044	0.061	0.031
19 – 34	0.279	0.626	0.596	0.525
35 – 49	0.237	0.222	0.271	0.337
> 49	0.182	0.108	0.071	0.107

Table B.2 Prior marginal probabilities for Age.

Education

	Freeler	Ilse	MSN	WOL
Counts	153	137	391	128
high school	0.325	0.276	0.286	0.296
College	0.202	0.167	0.192	0.203
high school graduate	0.124	0.145	0.146	0.132
Graduate school	0.052	0.073	0.054	0.086
College graduate	0.130	0.189	0.181	0.125
MBA	0.130	0.117	0.099	0.125
self-educated	0.033	0.030	0.041	0.032

Table B.3 Prior marginal probabilities for Education.

Position in the Household

Education	high school	college	high school graduate	graduate school	college graduate	MBA	self-educated
Counts	49	30	18	7	19	19	5
breadwinner	0.603	0.601	0.758	0.305	0.725	0.576	0.81
partner of	0.091	0.145	0.048	0.132	0.181	0.188	0.178

breadwinner							
child of breadwinner	0.254	0.254	0.106	0.559	0.051	0.105	0.007
other	0.052	0.001	0.088	0.004	0.043	0.132	0.006

Table B.4 Conditional probabilities for Position_Household for Freeler.

Age	< 19	19-34	35-49	> 49
Counts	6	83	27	12
breadwinner	0.170	0.541	0.884	0.658
partner of breadwinner	0.009	0.097	0.112	0.332
Child of breadwinner	0.810	0.265	0.002	0.005
Other	0.009	0.097	0.002	0.005

Table B.5 Conditional probabilities for Position_Household for Ilse.

Gender	male				female			
Age	< 19	19-34	35-49	> 49	< 19	19-34	35-49	> 49
Counts	18	158	67	17	6	54	31	9
breadwinner	0.167	0.550	0.850	0.878	0.168	0.296	0.483	0.442
partner of breadwinner	0.002	0.032	0.030	0.060	0.005	0.296	0.483	0.551
child of breadwinner	0.829	0.367	0.030	0.002	0.821	0.296	0.001	0.003
other	0.002	0.051	0.090	0.060	0.005	0.111	0.033	0.003

Table B.6 Conditional probabilities for Position_Household for MSN.

Gender	male	female
Counts	12	49
breadwinner	0.614	0.305
partner of breadwinner	0.046	0.547
child of breadwinner	0.268	0.124
other	0.072	0.023

Table B.7 Conditional probabilities for Position_Household for WOL.

Layout

Education	high school	college	high school graduate	graduate school	college graduate	MBA	self-educated
Counts	36	28	18	5	16	16	5
not clear	0.703	0.553	0.350	0.404	0.824	0.701	0.210
neutral	0.271	0.351	0.499	0.588	0.063	0.185	0.597
clear	0.026	0.096	0.151	0.009	0.113	0.114	0.193

Table B.8 Conditional probabilities for Layout for Freeler.

	poorly	good	very good
Counts	130	130	130
Layout	0.361	0.506	0.132

Table B.9 Marginal prior probabilities for Layout for Ilse.

Position_Household	breadwinner	partner of breadwinner	child of breadwinner	other
Counts	161	37	72	16
poorly	0.377	0.333	0.353	0.567
good	0.432	0.48	0.551	0.428
very good	0.191	0.187	0.097	0.0050

Table B.10 Conditional probabilities for Layout for MSN.

Education	high school	college	high school graduate	graduate school	college graduate	MBA	self-educated
Counts	27	22	15	8	14	9	4
not clear	0.259	0.362	0.069	0.128	0.287	0.330	0.253
neutral	0.181	0.176	0.321	0.244	0.276	0.005	0.247
clear	0.560	0.461	0.610	0.628	0.437	0.665	0.500

Table B.11 Conditional probabilities for Layout for WOL.

Look_Feel

Layout	not clear	neutral	clear
Counts	83	50	16
poorly	0.462	0.401	0.077
good	0.414	0.489	0.394
very good	0.124	0.11	0.528

Table B.12 Conditional probabilities for Look_Feel for Freeler.

Layout	not clear	neutral	clear
Counts	47	66	17
poorly	0.371	0.148	0.589
Good	0.563	0.670	0.420
very good	0.066	0.182	0.521

Table B.13 Conditional probabilities for Look_Feel for Ilse.

Layout	not clear	neutral	clear
Counts	115	151	51
Poorly	0.518	0.240	0.122
Good	0.432	0.611	0.372
very good	0.050	0.148	0.506

Table B.14 Conditional probabilities for Look_Feel for MSN.

Layout	not clear	neutral	clear
Counts	31	28	53
Poorly	0.260	0.498	0.115
Good	0.525	0.400	0.649
very good	0.215	0.102	0.236

Table B.15 Conditional probabilities for Look_Feel for WOL.

Stickiness

Navigation	poorly	good	v. good
Counts	64	46	41
< 53	0.286	0.314	0.138
53 - 99	0.259	0.183	0.276
99 - 196	0.180	0.185	0.330
> 196	0.274	0.318	0.256

Table B.16 Conditional probabilities for Stickiness for Freeler.

Navigation	poorly	good	v. good
Counts	31	26	14
< 79	0.198	0.271	0.291
79 - 157	0.224	0.152	0.422
157 - 258	0.238	0.321	0.080
> 258	0.340	0.255	0.207

Table B.17 Conditional probabilities for Stickiness for Ilse.

Navigation	poorly	good	v. good
Counts	110	112	86
< 48	0.227	0.287	0.190
48 - 117	0.241	0.271	0.255
117 - 211	0.280	0.249	0.249
< 211	0.253	0.193	0.305

Table B.18 Conditional probabilities for Stickiness for MSN.

Return	unlikely	likely	very likely
Counts	19	47	53
< 69	0.566	0.248	0.165
69 – 148	0.260	0.268	0.255
148 - 319	0.111	0.234	0.300
>319	0.063	0.250	0.280

Table B.19 Conditional probabilities for Stickiness for WOL.

Appendix C

Marginal probabilities

For marginal probabilities of Age, Gender, and Education, please refer to Appendix B.

	Freeler	Ilse	MSN	WOL
breadwinner	0.625	0.610	0.551	0.519
partner of breadwinner	0.126	0.123	0.122	0.201
child of breadwinner	0.197	0.203	0.264	0.223
other	0.051	0.064	0.064	0.057

Table C.1 Prior marginal probabilities for Position_Household.

Layout	Freeler	WOL	MSN	Ilse
not clear	0.612	0.256	0.377	0.361
neither clear nor not clear	0.305	0.196	0.469	0.506
clear	0.083	0.549	0.154	0.132

Table C.2 Prior marginal probabilities for Layout.

Look&Feel	Freeler	WOL	MSN	Ilse
negative	0.411	0.227	0.327	0.216
positive	0.435	0.569	0.507	0.599
very positive	0.154	0.204	0.166	0.185

Table C.3 Prior marginal probabilities for Look&Feel.

Navigation	Freeler	WOL	MSN	Ilse
poorly	0.414	0.384	0.343	0.435
somewhat	0.313	0.433	0.375	0.366
highly	0.273	0.183	0.281	0.199

Table C.4 Prior marginal probabilities for Navigation.

Freeler		WOL		MSN		Ilse	
< 53	0.255	< 69	0.262	< 48	0.239	< 79	0.243
53 - 99	0.240	69-148	0.261	48 - 117	0.256	79 - 157	0.237
99 - 196	0.222	148-319	0.243	117-211	0.259	157- 258	0.237
> 196	0.283	>319	0.234	< 211	0.245	> 258	0.283

Table C.5 Prior marginal probabilities for Stickiness.

Appendix D

1) Conditional probabilities for model in Chapter 5 - Measurement relationships for Model 1 with all indicators.

Trust	low	mod	high
Counts	187.6	61.9	166.4
low	0.669	0.038	0.021
moderate	0.307	0.721	0.216
high	0.024	0.241	0.762

Table D.1 Conditional probabilities for the indicator Tr1 given the latent construct Trust.

Trust	low	mod	high
Counts	187.6	61.9	166.4
low	0.398	0.009	0.005
moderate	0.445	0.532	0.040
high	0.156	0.458	0.954

Table D.2 Conditional probabilities for the indicator Tr2 given the latent construct Trust.

Trust	low	mod	high
Counts	187.6	61.9	166.4
low	0.627	0.049	0.012
moderate	0.294	0.707	0.181
high	0.077	0.243	0.807

Table D.3 Conditional probabilities for the indicator Tr3 given the latent construct Trust.

Trust	low	mod	high
Counts	187.6	61.9	166.4
low	0.441	0.035	0.011
moderate	0.383	0.631	0.097
high	0.174	0.332	0.891

Table D.4 Conditional probabilities for the indicator Tr4 given the latent construct Trust.

Involvement	low	mod	high
Counts	83.6	118.9	213.4
low	0.833	0.129	0.042
moderate	0.136	0.795	0.283
high	0.029	0.075	0.675

Table D.5 Conditional probabilities for the indicator Inv1 given the latent construct Involvement.

Involvement	low	mod	high
Counts	83.6	118.9	213.4
low	0.782	0.062	0.029
moderate	0.161	0.800	0.137
high	0.056	0.137	0.833

Table D.6 Conditional probabilities for the indicator Inv2 given the latent construct Involvement.

Involvement	low	mod	high
Counts	83.6	118.9	213.4
low	0.808	0.037	0.031
moderate	0.137	0.893	0.061
high	0.054	0.069	0.906

Table D.7 Conditional probabilities for the indicator Inv3 given the latent construct Involvement.

Involvement	low	mod	high
Counts	83.6	118.9	213.4
low	0.875	0.098	0.048
moderate	0.102	0.818	0.225
high	0.022	0.083	0.726

Table D.8 Conditional probabilities for the indicator Inv4 given the latent construct Involvement.

Involvement	low	mod	high
Counts	83.6	118.9	213.4
low	0.753	0.073	0.030
moderate	0.190	0.749	0.118
high	0.055	0.176	0.850

Table D.9 Conditional probabilities for the indicator Inv5 given the latent construct Involvement.

2) Model with three indicators for each construct model in Chapter 5

Marginal probabilities

	Satisfaction	Trust	Involvement	Loyalty
low	0.056	0.098	0.274	0.075
moderate	0.412	0.421	0.489	0.389
high	0.531	0.479	0.235	0.534

Table D.10 Prior marginal probabilities.

Structural model

Satisfaction	low	mod	high
Counts	21.7	171.9	222.4
low	0.622	0.080	0.057
moderate	0.324	0.693	0.220
high	0.052	0.225	0.721

Table D.11 Conditional probabilities for Trust given Satisfaction.

Satisfaction	low			mod			high		
Trust	low	mod	high	low	mod	high	low	mod	high
Counts	5.845	15.8	0.001	109.5	29.9	32.4	72.2	16.1	134.0
low	0.779	0.456	0.231	0.898	0.379	0.106	0.909	0.185	0.108
moderate	0.206	0.516	0.620	0.086	0.589	0.564	0.061	0.720	0.419
high	0.015	0.028	0.150	0.016	0.032	0.331	0.030	0.095	0.473

Table D.12 Conditional probabilites for Involvement given Satisfaction and Trust.

Involvement	low			mod			high		
Trust	low	mod	high	low	mod	high	low	mod	high
Counts	3.6	1.0	80.0	132.5	5.5	75.4	51.4	56.5	11.0
low	0.559	0.046	0.105	0.270	0.024	0.010	0.314	0.044	0.014
moderate	0.227	0.758	0.333	0.431	0.587	0.292	0.271	0.487	0.074
High	0.215	0.196	0.563	0.299	0.389	0.698	0.415	0.469	0.911

Table D.13 Conditional probabilites for Loyalty given Trust and Involvement.

Measurement models

Trust	low	mod	high
Counts	187.6	61.9	166.4
low	0.450	0.038	0.003
moderate	0.354	0.591	0.051
high	0.195	0.371	0.946

Table D.14 Conditional probabilities for the indicator Tr2 given the latent construct Trust.

Trust	low	mod	high
Counts	187.6	61.9	166.4
low	0.760	0.074	0.015
moderate	0.162	0.728	0.204
high	0.077	0.197	0.780

Table D.15 Conditional probabilities for the indicator Tr3 given the latent construct Trust.

Trust	low	mod	high
Counts	187.6	61.9	166.4
low	0.619	0.039	0.008
moderate	0.271	0.657	0.134
high	0.109	0.303	0.856

Table D.16 Conditional probabilities for the indicator Tr4 given the latent construct Trust.

Involvement	low	mod	high
Counts	83.6	213.4	118.9
low	0.837	0.048	0.014
moderate	0.124	0.886	0.074
high	0.038	0.065	0.911

Table D.17 Conditional probabilities for the indicator Inv3 given the latent construct Involvement.

Involvement	low	mod	high
Counts	83.6	213.4	118.9
low	0.936	0.090	0.032
moderate	0.052	0.836	0.224
high	0.012	0.073	0.743

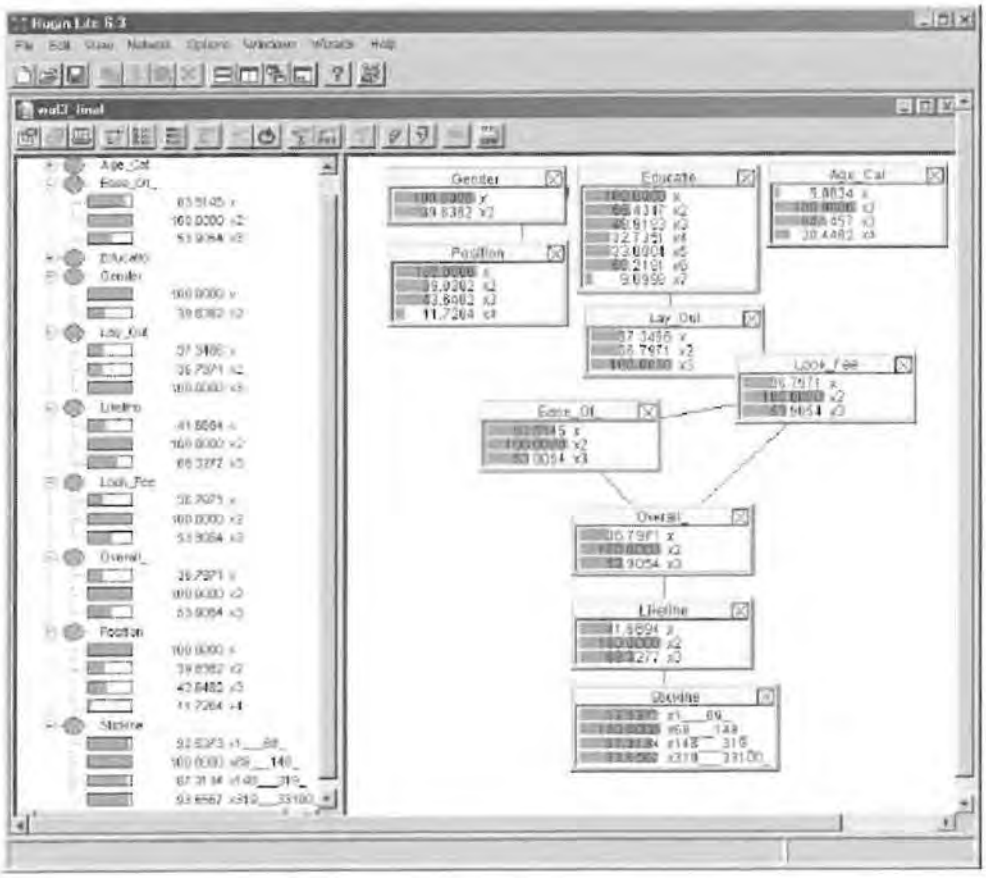
Table D.18 Conditional probabilities for the indicator Inv4 given the latent construct Involvement.

Involvement	low	mod	high
Counts	83.6	213.4	118.9
low	0.794	0.069	0.030
moderate	0.131	0.762	0.140
high	0.074	0.168	0.830

Table D.19 Conditional probabilities for the indicator Inv5 given the latent construct Involvement.

Appendix E

Most probable configuration for the profile of a visitor of the WOL website (see Section 4.7.7.)



The states that belong to the most probable configuration are those with the probability of 1 (here 100).

Bibliography

Agresti, A., 1984, Analysis of ordinal categorical data. New York: Wiley.

Akaike, H., 1974. "A new look at the statistical model identification". *IEEE Trans. Autom. Contr.* Vol. 19, pp. 716-723.

Aladwani, A., and P. Palvia, 2002, "Developing and validating an instrument for measuring user-perceived web quality". *Information & Management*, Vol. 39, pp. 467–476.

Alexander, C., 2000, "Bayesian Methods for Measuring Operational Risks," ISMA Centre Research Reports, Reading University, United Kingdom.

Allen, Natalie J., and John P. Meyer, 1990, "The Measurement and Antecedents of Affective, Continuance and Normative Commitment to the Organization", *Journal of Occupational Psychology*, Vol. 63, pp. 1-18.

Allen, Natalie J., and John P. Meyer, 1991, "A three-component conceptualization of organizational commitment". *Human resource management review*, Vol. 1, pp. 61–89.

Allen, Natalie J., and John P. Meyer, 1996, "Affective, Continuance, and Normative Commitment to the Organization: An Examination of Construct Validity", *Journal of Vocational Behavior*, Vol. 49, pp. 252-276.

Allen, D. and M. Willburn, 2002, Linking Customer and Employee Satisfaction to the Bottom Line. ASQ Quality Press, Milwaukee, Wisconsin.

Allen, D., 2004, Customer Satisfaction Research Management, ASQ Quality Press.

Andersen, S. K., K. G. Olesen, F. V. Jensen, and F. Jensen. 1989, "HUGIN - a shell for building Bayesian belief universes for expert systems". In: *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pp. 1080-1085, Detroit, Michigan, Aug. 20-25 1989.

Anderson, James C. and David W. Gerbing, 1982. "Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement," *Journal of Marketing Research*, Vol. 19 (November), pp. 453-60.

Anderson, James C. and David W. Gerbing (1988), "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin*, Vol. 103, No. 3, pp. 411-423.

Anderson, E. and B.A. Weitz, 1992, "The use of pledges to build and sustain commitment in distribution channels", *Journal of Marketing Research*, Vol. 29 (February), pp. 85-97.

Anderson, E.W. and Mittal, V., 2000, "Strengthening the satisfaction-profit-chain.", *Journal of Service Research*, Vol. 3, No. 2, pp. 107-120.

Anderson, E. W. and M. W. Sullivan. 1993. "The Antecedent and Consequences of Customer Satisfaction for Firms", *Marketing Science*, Vol. 12 (Spring), pp. 125-143.

Anderson, R., and R. Lenz, 2001, "Modelling the Impact of Organizational Change: A Bayesian Network Approach," *Organizational Research Methods*, Vol. 4, No. 2 (April), pp. 112-130.

Anderson, R. E., and S. Srinivasan, 2003, "E-Satisfaction and E-Loyalty: A Contingency Framework", *Psychology and Marketing*, Vol. 20, No. 2, pp. 123–138.

Andreassen, S., Finn V. Jensen, Stig K. Andersen, Bjørn Falck, Uffe Kjærulff, Marianne Woldbye, Anders R. Sørensen, Annelise Rosenfalck & Frank Jensen. 1989. "MUNIN - an expert EMG assistant", In: Desmedt, J. E. (ed.), *Computer-Aided Electromyography and Expert Systems*, ch. 21, pp. 255-277, Elsevier Science Publishers B.V., Amsterdam.

Andreassen, S., Jensen, F. V. & Olesen, K. G., 1991, "Medical expert systems based on causal probabilistic networks". *International Journal of Bio-Medical Computing*, Vol. 28, pp. 1-30.

Andreassen, T., Lindestad, B., 1998, "Customer loyalty and complex services: The impact of corporate image on quality, customer satisfaction and loyalty for customers with varying degrees of service expertise", *International Journal of Service Industry Management*, Vol.9, No. 1, pp. 7-23.

Armstrong, J. S., R. Brodie, and A. Parsons, 2001, "Hypotheses in Marketing Science: Literature Review and Publication Audit", *Marketing Letters*, Vol. 12, No. 2, pp. 171-187.

Assael, H., 1992. Consumer behavior and marketing action. Boston, MA: PWS-KENT Publishing Company.

Babakus , Ferguson, and Jöreskog, 1987, "The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*. Vol. 24, pp. 222–228.

Bagozzi, R. P. and C. Fornell, 1982, "Theoretical Concepts, Measurement, and Meaning," In: C. Fornell (Ed.) A Second Generation of Mulivariate Analysis: Praeger, Vol. 2, pp. 5-23.

Bagozzi, R.P., 1984, "A Prospectus for Theory Construction in Marketing," *Journal of Marketing*, Vol. 48 (Winter), pp. 11-29.

Bagozzi, R., and Y. Yi, 1989, "The Degree of Intention Formation as a Moderator of the Attitude-Behavior Relation". *Social Psychology Quarterly*, Vol. 52 (December), pp. 266-279.

Bagozzi, R., 1994a, "Measurement in Marketing Research", In: R. Bagozzi (ed), Principles of Marketing Research, Oxford, UK: Blackwell, pp. 1-49.

Bagozzi, R., 1994b , "Structural Equation Models in Marketing Research: Basic Principles", In: R. Bagozzi (ed), Advanced Methods of Marketing Research. Oxford, UK: Blackwell, pp. 317-385.

Bagozzi, R., and Baumgartner, H., 1994, "The Evaluation of Structural Equation Models and Hypothesis Testing", In: R. Bagozzi (eds.), Principles of Marketing Research Cambridge, pp. 386-422.

Bagozzi, R..P., Y. Yi and K. Nassen, 1999. "Representation of Measurement Error in Marketing Variables: Review of approaches and extension to three-facet designs," *Journal of Econometrics*, Vol. 89, pp. 393-421.

Bang,* J.-W., A. Papas,* and D. Gillies, 2003, "Interpretation of Hidden Node Methodology with Network Accuracy." Technical Report: ISSN 1469-4166, Dept. of Computing, Imperial College, London, UK, 2003.

Bansal, H. S., G. McDougall, S. S. Dikolli, and K. L. Sedatole, 2003, "Relating E-satisfaction to Behavorial Outcomes: An Empirical Study", *Journal of Services Marketing*, Vol. .

Barclay, D.W., 1991. "Interdepartmental conflict buying: The impact of the organizational context". *Journal of Marketing Research*, Vol. 28, pp. 145-159.

Bauer, E., D. Koller, and Y. Singer. 1997, "Update rules for parameter estimation in Bayesian networks". In: *Proceedings of the 13th Annual Conference on Uncertainty in AI* (UAI), 1997.

Bauer, H., and M. Hammerschmidt, 2002, "Developing and Validating a Quality Assessment Scale For Web Portals", Research Paper for the WSEAS Conference on E-Commerce, Band, 2002.

Baumgartner, H., and Ch. Homburg, 1996, "Applications of Structural Equation Modeling in Marketing and Consumer Research: A Review." *International Journal of Research in Marketing*. Vol. 13, No. 2, pp. 139-161.

Bearden, William O. and Jesse E. Teel, 1983, "Selected Determinants of Consumer Satisfaction and Complaint Reports," *Journal of Marketing Research*, Vol. 20 (February), pp. 21-28.

Beatty, Sharon E., Lynn R. Kahle, Pamela Homer, 1988, "The Involvement- Commitment Model: Theory and Implications", *Journal of Business Research*, 16 (2), pp. 149-167.

Berry, L. L., 1995. "Relationship Marketing of Services – Growing Interest, Emerging Perspectives". *Journal of the Academy of Marketing Science*. Vol. 23 (Winter), pp. 235-245.

Bitner, Mary Jo. 1990, "Evaluating Service Encounters: The Effects of Physical Surroundings and Employee Responses", *Journal of Marketing*, Vol. 54 (April), pp. 69-82.

Bitner, Mary Jo and Amy R. Hubbert, 1994, "Encounter Satisfaction Versus Overall Satisfaction Versus Quality," In: *Service Quality: New Directions in Theory and Practice*, ed. Roland T. Rust and Richard L. Oliver, Thousand Oaks, CA: Sage Publications, Inc., pp. 72-94.

Bloch, Peter H., and Marsha L. Richins. 1983, "A Theoretical Model for the Study of Product Importance Perceptions", *Journal of Marketing*, Vol. 47 (Summer), pp. 69-81.

Bloch, P. H., & Richins, M. L. (1983). Shopping without purchase: An investigation of consumer browsing behavior. Advances in Consumer Research, pp. 389-393.

Blodgett, J. G., and R. D. Anderson, 2000, "A Bayesian Network Model of the Consumer Complaint Process," *Journal of Service Research*, Vol. 2, No. 4, pp. 321-338.

Bloemer, J.M.M., and Ko de Ruyter, 1995. "Integrating service quality and satisfaction: pain in the neck or marketing opportunity?", *Journal of consumer satisfaction, dissatisfaction, and complaining behavior*, Vol. 8, pp. 44-52.

Bloemer, J., and H. Kasper, 1995. "The complex relationship between consumer satisfaction and brand loyalty". *Journal of Economic Psychology*, Vol. 16, pp. 311-329.

Bloemer, J., Ko de Ruyter and Pascal Peeters, 1998, "Investigating Drivers of Bank Loyalty: the Complex Relationship Between Image, Service Quality And Satisfaction", *International Journal of Bank Marketing*, Vol. 16, No. 7, pp. 276-286.

Bloemer, J., and Ko de Ruyter, 1999, "Customer Loyalty in High and Low Involvement Service Settings: The Moderating Impact of Positive Emotions", *Journal of Marketing Management*, Vol. 15, pp. 315-330.

Bloemer, J., Gaby Odekerken-Schröder and Leen Kestens (2001 ?), "The Effect of Consumer Values and Consumer Relationship Proneness on Service Satisfaction, Commitment and Store Loyalty", ???.

Bollen, K. A., 1989, Structural equations with latent variables, Wiley-Interscience.

Bouckaert, R. R. 1993, "Probalistic Network Construction Using the Minimum Description Length Principle". *ECSQARU*, pp. 41-48.

Buchanan, B. G. and E. H. Shortliffe, 1984, Rule-Based Expert Systems, Addison-Wesley, Reading, MA, Menlo Park, CA.

Bolton, R.N, 1998, "A Dynamic Model of the Duration of the Customer Relationship with a Continuous Service Provider: The Role of Satisfaction", *Marketing Science*, Vol. 17, No. 1, pp. 45-65.

Bolton, R.N. and Drew J.H., 1991, "A longitudinal analysis of the impact of service changes on customer attitudes." , *Journal of Marketing*, Vol. 55, No. 1, pp. 1-10.

Brandt, D.R. (1988), "How service marketers can identify value-enhancing elements.", *The Journal of Services Marketing*, 2, (3), 35-41.

Brodie, R. J. and P. J. Danaher, 2000, "Building models for marketing decisions: Improving empirical procedures." *International Journal of Research in Marketing*, Vol. 17, pp. 135–139.

Buchanan, B. G., and E. H. Shortliffe. 1984. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison-Wesley, Reading, MA.

Bullock, H. E., L. L. Harlow, and S. A. Mulaik, 1994, "Causation Issues in Structural Equation Modeling Research," *Structured Equation Modeling*, Vol. 1, No. 3, pp. 253-267.

Buntine, W. L., 1991, "Theory Refinement on Bayesian Networks". UAI 1991, pp. 52-60.

Cadotte, Ernest R., Robert B. Woodruff and Roger L. Jenkins, 1987, "Expectations and Norms in Models of Consumer Satisfaction," *Journal of Marketing Research*, Vol. 24 (August), pp. 305-314.

Campbell, D.T., 1969, "Definitional versus multiple operationalism." *Et Al*, 2, 14-17.

Campbell, D.T., and Fiske, D. W., 1959. "Convergent and discriminant validation by the multi-trait multimethod matrix." *Psychological Bulletin*, Vol. 56, pp. 85-105.

Castillo, E., J.M. Gutierrez, and A.S. Hadi, 1995, "Parametric Structure of Probabilities in Bayesian Networks," In: C. Froidevaux and J. Kohlas (Eds.), Lecture Notes in Artificial Intelligence: Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Springer Verlag, New York, Vol. 946, pp. 89-98.

Castillo, E., J. M. Gutierrez, and A. S. Hadi, 1997. Expert Systems and Probabilistic Network Models. Springer.

Cheeseman, P., and J. Stutz, 1995. "Bayesian Classification (AutoClass): Theory and Results", In: Advances in Knowledge Discovery and Data Mining, Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, & Ramasamy Uthurusamy, Eds. The AAAI Press, Menlo Park.

Chen, J., 2003, "An Entropy Theory of Psychology and its Implication to Behavioral Finance".

Chen, Zhan (Sandy), Sharon A. DeVaney, and Sandra S. Liu, 2003, "Consumers' Value Perception of an E-Store and Its Impact on E-Store Loyalty Intention", in: Proceedings of the Seventh Triennial AMS/ACRA Retailing Conference 2003, November 6-9, Columbus OH.

Cheng, J. and Greiner, R., 1999, "Comparing Bayesian Network Classifiers". In: *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, UAI'99.

Chickering, D.M., and D. Heckerman, 1997, "Efficient Approximations for the Marginal Likelihood of Bayesian Networks With Hidden Variables". *Machine Learning*, Vol. 29, pp. 181-212.

Chickering, D.M., D. Geiger and D. Heckerman, 1995. "Learning Bayesian Networks: Search Methods and Experimental Results". In: *Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*, pp. 112-128.

Chin, W. W., 1998, "The Partial Least Squares Approach to Structural Equation Modeling," in G. A. Marcoulides (Ed.) Modern Methods for Business Research, London, pp. 295-336.

Churchill, G. A., and C. Surprenant, 1982, "An Investigation into the Determinants of Customer Satisfaction," *Journal of Marketing Research*, Vol. 19 (November), pp. 491-504.

Cook T.D., Campbel D.T., 1979. "Quasi-Experimentation. Design and Analysis Issues for Field Settings". Rand McNally College Publishing Company, Chicago.

Cooper, D. R. & Emory, C. W., 1995. Research methods. Homewood, IL: Richard D. Irwin, Inc.

Cooper, G., 1990, "Computational Complexity of Probabilstic Inference Using Bayesian Belief Networks (Research Note)," *Artificial Intelligence*, Vol. 42, No. 2/3, pp. 393-405.

Cooper, G., and E. Herskovits, 1992, "A Bayesian Method for the Induction of Probabilistic Networks from Data", *Machine Learning*, Vol. 9, pp. 309-347.

Copeland, M.T., 1923, "Relation of Consumers' Buying Habits to Marketing Methods", *Harvard Business Review*, Vol. 1 (April), pp. 282-289.

Coupe, V. M. H., & Van der Gaag, L. C. 1997. "Supporting Probability Elicitation by Sensitivity Analysis". Lecture notes in computer science. Knowledge acquisition, Modeling and Management, 1319, pp. 335-340.

Cox, J., and B. G. Dale, 2001, "Service Quality and eCommerce: An Exploratory Analysis". *Managing Service Quality* (Electronic), Vol. 11, No 2, pp. 121-131.

Cronin, J. Joseph, Jr, Steven A. Taylor, 1992, "Measuring Service Quality: A Reexamination and Extension", *Journal of Marketing*, Vol. 56 (July), pp. 55-68.

Cronin, J. Joseph, Jr, Michael K. Brady, and G. Tomas M. Hult (2000), "Assessing the Effects of Quality, Value, and Customer Satisfaction on Consumer Behavioral Intentions in Service Environments", *Journal of Retailing*, Vol. 76, No. 2, pp. 193-218.

Crosby, Lawerence A., Evans, Kenneth R, and Cowles, Deborah., 1990. "Relationship Quality in Services Selling: An Interpersonal Influence Perspective". *Journal of Marketing*. Vol. 54 (July), pp. 68-81.

Cunningham, R.M., 1956, "Brand Loyalty - What Where How Much?", *Harvard Business Review*, Vol. 34, pp. 116-128.

Cutler M., and J. Sterne, 2000, E-Metrics, Business Metrics for the New Economy, (NetGenesis, www.netgen.com/emetrics/).

D'Ambrosio, B., 1999. "Inference in Bayesian networks". *AI Magazine*. Summer 1999, pp. 21-36.

Darden, W. R., and J. K. L. Schwinghammer, 1985. "The influence of social characteristics on perceived quality in patronage choice behavior". In: Jacoby, J. and J. C. Olson. Perceived quality: how consumers view stores and merchandise. Lexington Books: Lexington.

Darwiche, A., 2003, "A Differential Approach to Inference in Bayesian Networks", *Journal of the ACM*, Vol. 50, No. 3 (May), pp. 280–305.

Dawid, A. P., 1992, "Applications of a general propagation algorithm for probabilistic expert systems". *Statistics and Computing*, Vol. 2, pp. 25-36.

Dawkins, P. M., & Reichheld, F. F., 1990, "Customer Retention as a Competitive Weapon." *Harvard Business Review*. Summer.

Day, G.S., 1970, Buyer Attitudes and Brand Choice Behavior, New York, The Free Press.

Day, G.S., 1969, "A Two Dimensional Concept of Brand Loyalty", *Journal of Advertising Research*, Vol. 9 (Sep.), pp. 29-35.

de Ruyter, Ko, José Bloemer, and Pascal Peeters, 1997, "Merging Service Quality and Service Satisfaction: An Empirical Test of an Integrative Model", *Journal of Economic Psychology*, Vol. 18, pp. 387-406.

de Ruyter, K., and J. Bloemer, 1999, "Customer Loyalty in Extended Service Settings: The Interaction Between Satisfaction, Value Attainment and Positive Mood," *Journal of Service Industry Management*, Vol. 10, No 3, pp. 320-336.

Dechter, R., 1996, "Bucket Elimination: A Unifying Framework for Probabilistic Inference" In: *Uncertainty in Artificial Intelligence*, UAI'1996, 1996, pp. 211-219.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society*, Series B, Vol. 39, pp. 1-38.

Dick, Alan S. and Kunal Basu, 1994, "Customer Loyalty: Toward an Integrated Conceptual Framework", *Journal of the Academy of Marketing Science*, Vol. 22, No. 2, pp. 99-113.

Dietterich, T., 1997, Machine-Learning Research. Four Current Directions", *AI Magazine*, Vol. 18, No. 4 (Winter), pp. 97-136.

Dillon, W. R., White, John B., Rao, Vithala R., Filak, Doug., 1997, "Good science", Marketing Research, Vol. 9, No. 4, pp. 22-32.

Doney, P.M. & J.P. Cannon, 1997. "An Examination of the Nature of Trust in the Buyer-Seller Relationship". *Journal of Marketing*, Vol. 51, pp. 35-51.

Donthu, N., 2001, "Does your web site measure up?", *Marketing Management*, November/December, pp. 29-32

Douglas, V., 1995, "Questionnaire too long? Try variable clustering." *Marketing News*, Vol. 29, No. 5, p. 38.

Doyle, P. & Fenwick, I, (1974-75) "How Store Image affects Shopping Habits in Grocery Chains", *Journal of Retailing*, Vol. 50, No. 4, pp. 39-52.

Duarte M., and Davies G., 2000, "Trust as a Mediating Variable in a Complex Model of Channel Member Behaviour", IMP 2000 Conference.

Duda, R., and Hart, P. 1973. Pattern Classification and Scene Analysis. John Wiley and Sons, Inc.

Dwyer, Robert F., Paul H. Schuur, and Sejo Oh, 1987, "Developing Buyer-Seller Relationships", *Journal of Marketing*, Vol. 51 (April), pp. 11-27.

Ehrenberg, A.S.C., 1972, Repeat Buying: Theory and Applications. Amsterdam: North Holland.

Ehrenberg, A.S.C. (1994), "Theory or Well Based Results: Which Comes First?" In: Laurient, G.; Lilien, G; and Pras, B (eds.) Research Traditions in Marketing, pp. 79-108. Kluwer.

Ehrenberg, Andrew S.C., Neil R. Barnard, Byron Sharp, 2000. "Decision models or descriptive models?" *International Journal of Research in Marketing*. Vol. 17, No. 2, pp. 147 – 158.

Elidan, G., N. Lotner, N. Friedman, and D. Koller. 2001, "Discovering hidden variables: A structure-based approach". In NIPS 13.

Elidan, G., and N. Friedman, 2001, "Learning the Dimensionality of Hidden Variables". In: *Proc. of the Seventeenth Conf. on Uncertainty in Artificial Intelligence (UAI)*, 2001.

Erevelles, S. and C. Leavitt, 1992. "Comparison of Current Models of Consumer Satisfaction/ Dissatisfaction", J*ournal of Consumer Satisfaction/Dissatisfaction and Complaining Behavior*, Vol. 5, pp. 104-114.

Feigl, H., 1970, "The orthodox' View of Theories: Remarks in Defence as Well as Critique." In: *Minnesota Studies in the Philosophy of science*, Vol. 4, M. Radner and S. Winkour (eds.), University of Minnesota Press, 3-16.

Fischer, J., and K. Kersting. 2003, "Scaled cgem: A fast accelerated EM". In: N. Lavrac, D. Gamberger, H. Blockeel, and L. Todorovski (ed.), *Proceedings of the Fourteenth European Conference on Machine Learning* (ECML-2003), pp. 133–144, Cavtat, Croatia, 2003.

Fornell, C., 1982. A Second Generation of Multivariate Analysis Methods, Vol. 1. Praeger, New York.

Fornell, C., 1992. "A national customer satisfaction barometer: The Swedish experience." *Journal of Marketing*, Vol. 56 (Jan.), pp. 6-21.

Fornell, C. and J. Cha, 1992. Partial least squares. Working paper, Graduate School of Business, University of Michigan, Ann Arbor, MI.

Fornell, C., & Yi, Y., 1992. "Assumptions of the two-step approach to latent variable modeling". *Sociological Methods & Research*, Vol. 20, No. 1, pp. 291-320.

Fornell, C. and J. Cha, 1994. Partial Least Squares. In: R. P. Bagozzi. (Ed.) Advanced Methods of Marketing Research. Oxford, Blackwell Publishers.

Fornell, C.R. and Lacker, D.F., 1981. "Two structural equation models with unobservable Variables and measurement error." *Journal of Marketing Research*, Vol. 18, pp. 39-50.

Fornell, C., M. Johnson, E. Anderson, J. Cha, B. Bryant, 1996, "The American Customer Satisfaction Index: Nature, Purpose, and Findings", *Journal of Marketing*, Vol. 60 (October), pp. 7-16.

Francis, J.E. and White, L., 2002, "PIRQUAL: a scale for measuring customer expectations and perceptions of quality in internet retailing," Proceedings of the Winter Educator's Conference, American Marketing Association, Chicago, IL, pp. 438-443.

Freedman, J., 1964. "Involvement, discrepancy, and change". *Journal of abnormal and Social Psychology*, Vol. 69, pp. 290-295.

Friedman, N., D. Geiger, M. Goldszmidt, 1997, "Bayesian Network Classifiers". *Machine Learning*, Vol. 29, pp. 131–163.

Friedman, N., 1998, "The Bayesian Structural EM Algorithm", In: *Proceedings of the 14th Conference on Uncertainty in AI*, 1998, pp. 129-138. San Francisco, CA: Morgan Kaufmann.

Furse, David H., Girish N. Punj and David E. Stewart, 1984, "A Typology of Individual Search Strategies Among Purchases of New Automobiles," *Journal of Consumer Research*, Vol. 10 (March), pp. 417-431.

Gaag, L. C. van der, and S. Renooij, 2001, "Evaluation Scores for Probabilistic Networks," In: B. Kröse, M. de Rijke, G. Schreiber and M. van Someren (Eds), In: *Proceedings of the Thirteenth Belgium-Netherlands Conference on Artificial Intelligence*, Amsterdam, pp. 109-116.

Gaag, L.C. van der, Renooij, S., Witteman, C.L.M., Aleman, B. M. P., & Taal, B.G., 2002. "Probabilities for a probabilistic network: A case-study in oesophageal cancer". *Artificial Intelligence in Medicine*, Vol. 25, No. 2, pp. 123-148.

Garbarino, E., and M. S. Johnson, 1999. "The Different Roles of Satisfaction, Trust, and Commitment in Customer Relationships", *Journal of Marketing*, Vol. 63, No. 2, pp. 70-87.

Gefen, D., Straub, D., and Boudreau, M.-C. 2000, "Structural Equation Modeling and Regression: Guidelines for Research Practice," *Communications of the Associations for Information Systems*, Vol. 4, No.7.

Geiger, D., Verma, T., and Pearl, J., 1990, "Identifying independence in Bayesian Networks". *Networks*, Vol. 20, pp. 507-533.

Geisser, S., 1975. "The predictive sample reuse method with applications". *Journal of the American Statistical Association*, Vol. 70, pp. 320-328.

Gerbing, D. W. and J. C. Anderson, 1988, "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," *Journal of Marketing Research*, Vol. 25 (May), pp. 186-192.

Geyskens, I., and J.-B. Steenkamp, 2000, "Economic and social satisfaction: Measurement and relevance to marketing channel relationships". *Journal of Retailing*, Vol. 76, No. 1, pp. 11-32.

Geyskens, I., J.-B. E.M. Steenkamp, Lisa K. Scheer, Nirmalya Kumar, 1996, "The Effects of Trust and Interdependence on Relationship Commitment: A Trans-Atlantic Study", *International Journal of Research in Marketing*, Vol. 13, pp. 303-317.

Giese, J. L., and J. A. Cote. 2000. "Defining Consumer Satisfaction", Academy of Marketing Science Review [Online] 00 (01), Available: http://www.amsreview.org/amsrev/theory/giese01-00.html.

Gommans, M., K.S. Krishnan, and K. B. Scheffold, 2001, "From Brand Loyalty to E-Loyalty: A Conceptual Framework", *Journal of Economic and Social Research*, Vol. 3, No. 1, pp. 43-58.

Good, I., 1961, "A causal calculus." *British Journal for the History of Science*, 11, 305–318.

Good, I. J., 1985, "Weight of evidence: a brief survey," In: J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (eds.). Bayesian Statistics 2, Proceedings of the Second Valencia International Meeting, September, 1983,. New York: North Holland.

Goodman, L. A., 1974. "Exploratory latent structure analysis using both identifiable and unidentifiable models". *Biometrika*, Vol. 61, pp. 215-231.

Greenleaf, E. A., 1992, "Improving rating scale measures by detecting and correcting bias components in some response styles." *Journal of Marketing Research*, Vol. 29 (May), p. 176.

Grönross, Ch., F. Heinonen, K. Isoniemi, and M. Lindholm, 2000, "The NEToffer model: a case example from virtual marketspace", *Management Decision*, Vol. 38, No. 4, pp. 243-252.

Gruen, Thomas W., John O. Summers, and Frank Acito, 2000, "Relationship Marketing Activities, Commitment, and Membership Behaviors in Professional Associations", *Journal of Marketing*, Vol. 64 (July), pp. 34-49.

Grunert, Suzanne C., and Hans Jorn Juhl, 1995, "Values, Environmental Attitudes, and Buying of Organic Food", *Journal of Economic Psychology*, Vol. 16, pp. 39-62.

Grunet, K.G. 1989, "Attributes, attribute value and their characteristics: a unifying approach and an example involving a complex household investment". *Journal of economic psychology*, Vol. 10, pp. 229-251.

Gundlach, Gregory T., Ravi S. Achrol and John T. Mentzer, 1995, "The Structure of Commitment in Exchange", *Journal of Marketing*, Vol. 59 (January), pp. 78-92.

Hagenaars, J. A., 1988. "Latent structure models with direct effects between indicators: local dependence models". *Sociological Methods and Research*, Vol. 16, pp. 379-405.

Hagenaars J., and R. Luijkx, 1990, "LCAG; Latent Class Models and Other Loglinear Models With Latent Variables, Manual LCAG, revised edition", Working Paper Series, Vol. 17, Tilburg University, Department of Sociology.

Hagenaars, J.A., 1988, "LCAG - Loglinear modelling with latent variables; a modified LISREL approach", In: W.E. Saris and I.N. Gallhofer (eds.), Sociometric Research, Vol. 2 Data Analysis, pp. 111-130. McMillan, London.

Hanson, N. R., 1963, The Concept of Positivism, A Philosophical Analysis. Cambridge: Cambridge University Press, 1963.

Heckerman, D. 1986. "Probabilistic Interpretation for MYCIN's Certainty Factors". In: Uncertainty in Artificial Intelligence, ed. Kanal, L. N. and J. F. Lemmer, pp. 167-196. Amsterdam, London, New York: Elsevier/North Holland.

Heckerman, D.E. and B.N. Nathwani, 1992, "Toward normative expert systems: Part II Probability-based representations for efficient knowledge acquisition and inference," *Methods of Information in Medicine*, Vol. 31, No. 2, pp. 106-116.

Heckerman. D., 1999, "A Tutorial on Learning with Bayesian Networks". In: Learning in Graphical Models, M. Jordan, ed.. MIT Press, Cambridge, MA.

Heckerman, D. E., Horvitz, E. J., & Nathwani, B. N., 1992, "Toward Normative Expert Systems: Part I The Pathfinder Project". KSL report 92-66.

Heckerman, D. and J. Breese. 1994, "A new look at causal independence". Technical Report MSR-TR-94-08, Microsoft Research, March.

Heckerman, D., J. Breese, K. Rommelse. 1995, "Decision-theoretic troubleshooting". *Communications of the ACM*, Vol. 38, pp. 49-57.

Heckerman, D., D. Geiger, D. Chickering. 1995, "Learning Bayesian networks: The Combination of Knowledge and Statistical Data". *Machine Learning*, Vol. 20, pp. 197-243.

Heckerman, D., C. Meek, and G. Cooper, 1999, "A Bayesian Approach to Causal Discovery". In: C. Glymour and G. Cooper (ed.), *Computation, Causation, and Discovery*, pp. 141-165. MIT Press, Cambridge, MA.

Hempel, C., 1965, "Aspects of Scientific Explanation" In: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York, The Free Press, pp. 331-496.

Henrion, M., 1988, "Propagation of uncertainty by logic sampling in Bayesian networks." In: *Uncertainty in Artificial Intelligence* 2 (J.F. Lemer, L.N. Kanal, eds.) Elsevier, Amsterdam, pp. 149-164.

Hill, N., and J. Alexander, 2000, Handbook of Customer Satisfaction and Loyalty Measurement. Gower Publishing Limited.

Hoffman, D.L. and T.P. Novak, 1996, "Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations," *Journal of Marketing*, Vol. 60 (July), pp. 50-68.

Hoffman, D.L., T.P. Novak, and M.A. Peralta, 1999, "Building Consumer Trust in Online Environments: The Case for Information Privacy," *Communications of the ACM*, Vol. 42, No. 4, pp. 80-85.

Hoque A.Y., Lohse G.L., 1999, "An Information Search Cost Perspective for Designing Interfaces for Electronic Commerce", *Journal of Marketing Research*, Vol. 36 (August), pp. 387-394.

Hsu, W. H., H. Guo, B. B. Perry, and J. A. Stilson. 2002, "A Permutation Genetic Algorithm for Variable Ordering in Learning Bayesian Networks from Data". In: *Proceedings of the Genetic and Evolutionary Computation Conference* (GECCO-2002), New York, NY.

Hulland, J.S. and D.N. Kleinmuntz, 1994. "Factors influencing the use of internal summary evaluations versus external information in choice." *Journal of Behavioral Decision Making*, 7 (2), 79-102.

Hulland, John, Yiu Ho Chow and Shunyin Lam, 1996, "Use of causal models in marketing research: A review", *International Journal of Research in Marketing*, Vol. 13, No. 2, April.

Hunt, S., 1976, "The Nature and Scope of Marketing", *Journal of Marketing*, Vol. 40 (July), pp. 17-28.

Hunt 1983, "General Theories and the Fundamental Explananda of Marketing", *Journal of Marketing*, Vol. 47 (Fall), pp. 9-17.

Hunt, Shelby D., 1991, *Modern Marketing Theory: Critical Issues in the Philosophy of Marketing Science*, South-Western Publishing Company, Cincinnati.

Iverson, Roderick D., and Donna M. Buttigieg, 1999, "Affective, Normative, and Continuance Commitment: Can the 'Right Kind' of Commitment Be Managed?", *Journal of Management Studies*, Vol. 36, No. 3 (May), pp. 307-333.

Jacoby, J., 1971. "A model of multi-brand loyalty". *Journal of Advertising Research*, Vol. 11, pp. 25-30.

Jacoby, J., and R. W. Chestnut, 1978, Brand Loyalty Measurement and Management, John Wiley and Sons.

Jensen, F., Lauritzen, S., Olesen, K. 1990. "Bayesian updating in recursive graphical models by local computations." *Computational Statisticals Quarterly*, Vol. 4, pp. 269-282.

Jensen, F. V., 1996. An Introduction to Bayesian Networks. UCL Press.

Jensen, F., 2001. Bayesian Networks and Decision Graphs. Springer.

Jensen, F. V., U. Kjærul, B. Kristiansen, H. Langseth, C. Skaanning, J. Vomlel, and M. Vomlelov´a, 2001, "The SACSO methodology for troubleshooting complex systems." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 15:321–333, 2001.

Johnson, M.D. and C. Fornell, 1987. "The nature and methodological implications of the cognitive representation of products." *Journal of Consumer Research*, 14 (Sep.), 214-228.

Jöreskog, K. G., and Sörbom, D., 1993, *LISREL 8 user's reference guide*. Chicago, IL: Scientific Software International.

Kaplan, A., 1964, The Conduct of Inquiry. Scranton, PA: The Chandler Publishing Company.

Kaynama, S.A. and Black, C.I. (2000), "A proposal to assess the service quality of online travel agencies: an exploratory study", *Journal of Professional Services Marketing*, Vol. 21, pp. 63-88.

Keaveney, S. M., and K. A. Hunt (1992), "Conceptualization and Operationalization of Retail Store Image: A Case of Rival Middle-Level Theories", *Journal of the Academy of Marketing Science*, Vol. 20, No. 2, pp. 165-175.

Kenny, D. A., and Judd, C. M., 1984. „Estimating the nonlinear and interactive effects of latent variables". *Psychological Bulletin*, Vol. 96, pp. 201–210.

Kipersztok, O., and H. Wang, 2001, "Another Look at Sensitivity of Bayesian Networks to Imprecise Probabilities". AISTATS 2001.

Kumar, N., L. K. Scheer, and J.-B. E.M. Steenkamp, 1995, "The effects of Supplier Fairness on Vulnerable Resellers", *Journal of Marketing Research*, Vol. 32 (February), pp. 54-65.

Kolmogorov, A. N., 1933, Foundations of the Theory of Probability. Chelsea, New York, 1950 (originally published in 1933 as Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer, Berlin).

Korgaonkar, P.K., Lund, D. and Price, B. 1985, "A structural equations approach toward examination of store attitude and store patronage behavior," *Journal of Retailing*, Vol. 61, No. 2, pp. 39-60.

Kumar, N., Hibbard, J.D. and Stern, L.W., 1994. "The Nature and Consequences of Marketing Channel Intermediary Commitment". MSI Working Paper, Report Nr. 94/115.

Kumar, Nirmalya, Lisa K. Scheer, and Jan-Benedict E.M. Steenkamp (1995), "The effects of Supplier Fairness on Vulnerable Resellers", *Journal of Marketing Research*, Vol. 32 (February), pp. 54-65.

Kunkel, J.H. & Berry, L.L., 1968, "A Behavioural Conception of Retail Image", *Journal of Marketing*, Vol. 32, pp. 21-27.

La Cerra, P., and R. Bingham, 2002, The Origin of Minds, New York: Harmony Books.

Lam, W., and F. Bacchus, 1993, "Learning Bayesian Belief Networks: An Approach Based on the {MDL} Principle", *Computational Intelligence*, Vol. 10, No. 4.

Larrañaga, P., C. Kuijpers, R. Murga, and Y. Yurramendi. 1996, "Learning Bayesian network structures by searching for the best ordering with genetic algorithms". *IEEE Transactions on System, Man and Cybernetics*. Vol 26., No. 4, pp. 487-493.

Laurent, G. and J.N. Kapferer, 1985. "Measuring consumer involvement profiles". *Journal of Marketing Research*, Vol. 22, pp. 41-53.

Lauritzen, S., 1995, "The EM algorithm for graphical association models with missing data". *Computational Statistics & Data Analysis*, Vol. 19, pp. 191-201.

Lauritzen, S. and Spiegelhalter, D., 1988. "Local computations with probabilities on graphical structures and their application to expert systems". *Journal of Royal Statistical Society B*, Vol. 50, pp. 157-224.

Lazer, W., 1962, "The role of models in marketing", *Journal of Marketing*, Vol. 26 (April), pp. 9-14.

Leeflang, P.S.H., Wittink, D.R., Wedel, M., Naert, P.A., 2000. Building Models for Marketing Decisions. Kluwer Academic Publishers, Dordrecht, Netherlands.

Leeflang, Peter S.H., Dick R. Wittink, 2000, "Building models for marketing decisions: Past, present and future". *International Journal of Research in Marketing*, Vol. 17, Nos. 2-3, pp. 105-126.

Levitt, T. 1983. The Marketing Imagination. New York: The Free Press.

Lilien, G. L., P. Kotler, K. S. Moorthy, 1992. Marketing Models. Englewood Cliffs, Prentice-Hall.

Lilien, Gary L. and Arvind Rangaswamy, 2000, "Modeled to Bits: Decision Models for the Digital, Networked Economy," *International Journal of Research in Marketing*, Vol. 17, Nos. 2-3, pp. 227-235.

Liljander, V., A.C.R. van Riel, and M. Pura, 2002, "Customer Satisfaction with E-Services: The Case Of An Online Recruitment Portal", In: (Eds.) Bruhn. M. and B. Stauss. The Yearbook of Services Management 2002 – E-Services.

Little, John D.C., 1994. "Modeling market response in large customer panels". In: Blattberg, Robert C., Rashi Glazer, and John D.C. Little (eds.). The marketing information revolution. Harvard Business School Press. Boston, 1994.

Little, R.J.A., and Rubin, D. B., 1987. Statistical Analysis with Missing Data. New York: John Wiley.

Liu, Chang, and Kirk P. Arnett, 2000, "Exploring the factors associated with wed site success in the context of electronic commerce", *Information & Management*, Vol. 38, pp 23-33.

Loiacono, Eleanor T., Watson, Richard T. & Goodhue, Dale L. 2000. WebQualTM: A measure of website quality. [Online]. Available WWW: http://www.terry.uga.edu/cisl/includes/pdf/webqual.pdf. [Accessed 6 December 2002].

Loiacono, Eleanor T., Chen, Daniel Q. & Goodhue, Dale L. 2002. WebQualTM revisited: predicting the intent to reuse a website. Eighth Americas Conference on Information Systems on the World Wide Web. [Online]. Available WWW: http://www.terry.uga.edu/people/dchen/AMCIS2002.pdf. [Accessed 25 March 2002].

Lord, F. M., and M. R. Novick, 1968, *Statistical Theories of Mental Test Scores*. Reading, MA. Addison-Wesley.

Lynch, Patrick D., Robert J. Kent and Srini S. Srinivasan, 2001, "The Global Internet Shopper: Evidence from Shopping Tasks in Twelve Countries," *Journal of Advertising Research*, Vol. 41, No. 3, pp. 15-23.

Lynch, J., and Ariely, D., 2000, "Wine on-line : Search costs and competition on price, quality and distribution", *Marketing science*, Vol. 19, No. 1, pp. 83-103.

Macintosh, Gerrard, and Lawrence S. Lockshin, 1997, "Retail relationships and store loyalty: A multi level perspective", *International Journal of Research in Marketing*, Vol. 14, pp. 487-497.

Madigan, D., Mosurski, K., and Almond, R. G., 1997, "Graphical Explanation in Belief Networks," *Journal of Computational and Graphical Statistics*, Vol. 6, No. 2.

Mahajan, Vijay and R. Venkatesh (2000), "Marketing Modeling for e-Business," *International Journal of Research in Marketing*, 17(2–3), pp. 215–225.

Malhotra, Naresh K., 1993. Marketing Research: An applied orientation. Prentice-Hall International Editions.

Martilla, John and James, 1977, "Importance-Performance Analysis" *Journal of Marketing*, Vol. 41, No. 1, pp. 77-79.

Mazursky D., and Jacoby J., 1986, "Exploring the Development of Store Images,", *Journal of Retailing*, Vol. 62, pp. 145-165.

McClendon, M., and D. O'Brien, 1988, "Question Order Effects on the Determinants of Subjective Well-Being", *Public Opinion Quarterly*, Vol. 52, pp. 351-364.

McDonald, R.P., 1996. "Path analysis with composite variables". *Multivariate Behavioral Research*, Vol. 31, No. 2, pp. 239-270.

McGarry, E., 1936, "The Importance of Scientific Method in Advertising", *Journal of Marketing*, Vol. 1 (October), pp. 82-86.

McLachlan, G., and Peel, D., 2000, Finite Mixture Models, NY: Wiley Publications.

Miller, G., 1956, "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information," *Psychological Review*, Vol. 63, pp. 81-97.

Mittal, B. and M.S. Lee, 1988. "Separating brand choice involvement from producer involvement via consumer involvement profiles". *Advances in Consumer Research*, Vol. 15, pp. 36-40.

Mittal V., Kumar P., Tsiros M., 1999, "Attribute – level performance, satisfaction, and behavioural intentions over time: A consumption-system approach", *Journal of Marketing*, Vol. 63, No. 2, pp. 88-101.

Mittal, V., W. Ross, and P. Baldasare, 1998, "The Asymmetric Impact of Negative and Positive Attribute-Level Performance on Overall Satisfaction and Repurchase Intentions", *Journal of Marketing*, Vol. 62, No. 1, pp. 33-47.

Mobius, 2002. "Concern with Customer Service...or Lack Thereof". eMarketer Daily, Issue 23.

Montoya-Weiss, M., Mueller, J., & Story, M., 2000. "Measuring universal design." Proceedings, Designing for the 21st Century II: An International Conference on Universal Design. Boston: Adaptive Environments Center.

Moorman, C., Deshpande, R., and Zaltman, G. (1993). "Factors Affecting Trust in Market Research Relationships". *Journal of Marketing* (January).

Morgan, M. G., and Henrion, M. 1990. *Uncertainty, A guide to dealing with uncertainty in qualitative risk and policy analysis*. Cambridge: Cambridge University Press.

Morgan, Robert M. and Shelby Hunt (1994), "The Commitment-Trust Theory of Relationship Marketing", *Journal of Marketing*, Vol. 58 (July), pp. 20-38.

Naert, and P. Leeflang, 1978, Building Implementable Marketing Models. Leiden, Nijhoff.

Naumann, E., and K. Giel, 1995, Customer Satisfaction Measurement And Management. Thomson Executive Press.

Neapolitan, R., 2004, Learning Bayesian Networks. Prentice Hall Series in AI.

Neter, J., W. Wasserman, and M. H. Kutner, 1990. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Design, 3rd edition. Boston, MA: Irwin.

Novak, T., Hoffman, D. and Yiu-Fai Yung, 2000, "Measuring the Customer Experience in Online Environments: A Structural Modeling Approach", *Marketing Science*, Vol. 19, No. 1.

Nunnally, J. C., 1967, Psychometric theory. New York: McGraw Hill.

Nunnally, J. C., Bernstein, I. H., 1994, Psychometric theory. New York.

O'Connor, J., and E. Galvin, 2001, Marketing in the digital age. Harlow : Prentice Hall.

Olesen, K. G., Kjærulff, U., Jensen, F., Jensen, F. V., Falck, B., Andreassen, S. and Andersen, S. K. (1989): A MUNIN Network for the Median Nerve - A Case Study on Loops. *Applied Artificial Intelligence*, Vol. 3, pp. 385-404.

Oliva, Terence A., Richard L. Oliver and Ian C. MacMillan, 1992, "A Catastrophe Model for Developing Service Satisfaction Strategies," *Journal of Marketing*, Vol. 56 (July), pp. 83-95.

Oliver, R.L. 1980. "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions". *Journal of Marketing Research*. Vol. 17, pp. 460-469.

Oliver, R. L., 1981. "Measurement and Evaluation of Satisfaction Processes in Retail Settings," *Journal of Retailing*, Vol. 57 (Fall), pp. 25-48.

Oliver, R. L., 1989. "Processing of the satisfaction response in consumption: a suggested framework and research propositions". *Journal of Customer Satisfaction, Dissatisfaction and Complaining Behaviour*. Vol. 2, pp. 1-16.

Oliver, R., 1993, "Cognitive, Affective and Attribute Bases of the Satisfaction Response," *Journal of Consumer Research*, Vol. 20 (December), pp. 418-430.

Oliver, R., 1997, *Satisfaction: A Behavioral Perspective on the Consumer*, New York: Irwin/McGraw-Hill.

Oliver, R. L., 1999, "Whence Customer Loyalty," *Journal of Marketing*, Vol. 63 (Special Issue), pp. 33-44.

Oliver, R., and J. Swan, 1989, "Equity and Disconfirmation Perceptions as Influences on Merchant and Product Satisfaction," *Journal of Consumer Research*, Vol. 16, pp. 372-383.

Oliver, R. L. and W. S. De Sarbo, 1988. "Response determinants in satisfaction judgements." *Journal of Consumer Research*, Vol. 14 (March), pp. 495-507.

O'Loughlin, C. and Coenders, G. 2002 "Application of the European Customer Satisfaction Index to Postal Services. Structural Equation Models versus Partial Least Squares." Working Papers of the Department of Economics, University of Girona, Vol. 4, pp. 1-28.

Ortinau, 1989. "The use of importance-performance analysis for improving the quality of marketing education: interpreting faculty-course evaluations." *Journal of Marketing Education*, Vol. 11, pp. 78-86.

Osman, M.Z. 1993, "A Conceptual Model of Retail Image Influences on Loyalty Patronage Behavior", *The International Review of Retail, Distribution and Consumer Research*, Vol. 31, pp. 149-166.

Panofsky, H. A., and G. W. Brier, 1968, *Some Applications of Statistics to Meteorology*, The Pennsylvania State University, University Park, Pennsylvania.

Parasuraman, A., Valarie A. Zeithaml, Leonard L. Berry, 1985, "A Conceptual Model of Service Quality and Its Implications for Future Research", *Journal of Marketing*, Vol. 49 (Fall), pp. 41-50.

Parasuraman, A., Valarie A. Zeithaml, Leonard L. Berry, 1988, "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality", *Journal of Retailing*, Vol. 64, No. 1, pp. 12-37.

Parasuraman, A., Zeithaml, V.A. and Berry, L.L., 1994, "Alternative scales for measuring service quality: A comparative assessment based on psychometric and diagnostic criteria". *Journal of Marketing*, Vol. 70, pp. 201-230.

Patterson, Paul G., 1993, "Expectations and product performance as determinants of satisfaction with a high-involvement purchase", *Psychology and Marketing*, Vol. 10, No.5, pp. 449-465.

Patterson, P.G., Johnson, L.W. and Spreng, R., 1997, "Modeling the determinants of customer satisfaction for business-to-business professional services", *Journal of the Academy of Marketing Science*, Winter, pp. 4-17.

Pearl, J., 1988, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo (CA).

Pearl, J. and Verma, T., 1990. "A Formal Theory of Inductive Causation". Technical Report R- 55, Cognitive Systems Labratory, Computer Science Dept. UCLA.

Peter, J. Paul and Jerry C. Olson, 1996, Consumer Behavior and Marketing Strategy, The McGraw-Hill Companies.

Peterson, Robert A., and William R. Wilson, 1992. "Measuring Customer satisfaction: Fact or artifact". *Journal of the Academy of Marketing Science*. Vol. 20, No. 1, pp. 61-71.

Ping, R.A., 1996, "Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients," *Multivariate Behavioral Research*, Vol. 31, No. 1, pp. 95-120.

Ranaweera, Ch., G. McDougall, and H. Bansal, 2004, "A Model of online Customer Behaviour: Moedrating Effects of Customer Characteristics." ??

Ranganathan, C. and S. Ganapathy, 2002, "Key dimensions of businessto-consumer web sites", *Information and Management*, Vol. 39, pp. 457-465.

Rangaswamy, A., 1993, "Marketing Decision Models: From Linear Programs to Knowledge-based Systems," In: J. Eliashberg and G. Lilien (Eds), Handbook of MS/OR in Marketing, Elsevier Publishing, pp. 733-771.

Ramoni, M., and P. Sebastiani, 1997, "Learning Bayesian Networks from Incomplete Databases", In: *Proceedings of the Thirteen Conference on Uncertainty in Artificial Intelligence*. Providence, Rhode Island, August 1-3 1997, pp. 401-408.

Ratchford, B. T., 1980, "The Value of Information for Selected Appliances," *Journal of Marketing Research*, Vol. 17, No. 1, pp. 14-25.

Render, B., and O'Connor, 1976, "The influence of price, store name, and brand name on perceptions of product quality", Journal of the Academy of Marketing Science, Vol. 4, pp. 722-730.

Reichheld, F., and W. Sasser, 1990, "Zero defections: Quality comes to to services", *Harvard Business Review*. Vol. 68 (September/October), pp. 105-11.

Reichheld, F.F. and Schefter, P. (2000), ``E-loyalty: your secret weapon on the Web", Harvard Business Review, July-August, pp. 105-13.

Richins, M., and P. Bloch, 1986, After the new wears off: The temporal context of product involvement", *Journal of consumer research*, Vol. 43, pp. 280 - 285.

Rigby, P., 1965, *Conceptual Foundations of Business Research*. New York, John Wiley and Sons.

Rigdon, E., 1998, Structural equation modeling. In: G. Marcoulides (ed.), *Modern methods for business research*. Mahwah, NJ: Lawrence Erlbaum, pp. 251-94.

Rigdon, E., R. E. Schumacker and W. Wothke, 1998, "A Comparative Review of Interaction and Nonlinear Modeling," In: R.E. Schumaker and G.A. Marcoulides (eds.), Interaction and Nonlinear Effects in Structural Equation Modeling, pp. 1-16. Mahwah, NJ: Lawrence Erlbaum.

Roberts, J. H., 2000, "The intersection of modelling potential and practice", *International Journal of Research in Marketing*, Vol. 17, pp. 127-134.

Robinson, R. W., 1977. „Counting unlabelled acyclic digraphs". Combinatorial Mathematics V, Lecture Notes in Mathematics 622, New York: Springer.

Rositer, J. R., 1994, "Commentary ...", In: Laurent, G., Lilien, G. L., and Pras, B., Research Traditions in Marketing, Kluwer, Boston.

Rossi, P., and G. Allenby, 2003, "Bayesian Statistics and Marketing," *Marketing Science*, forthcoming.

Rousseau, D. M., S. B. Sitkin, R. S. Burt, and C. Camerer. 1998. "Not So Different after All: A Cross-Discipline View of Trust." *Academy of Management Review*, Vol. 23, pp. 393-404.

Rudner, R., 1966, Philosophy of Social Science. Englewood Cliffs, NJ. Prentice-Hall, Inc.

Rundle-Thiele, K. and Lockshin, L., 2001, "Exploring The Dimensions of Loyalty", ANZMAC 2001 International Conference, Auckland, New Zealand (December).

Rusakov and Geiger, 2003, "Automated Analytic Asymptotic Evaluation of the Marginal Likelihood for Latent Models", In: Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence.

Russel, S., J. Binder, D. Koller,1 and K. Kanazawa, 1995, "Local learning in probabilistic networks with hidden variables". In: *Proceedings of the Fourteenth International Joint Conference on Articial Intelligence*, pp. 1146 -1152. Morgan Kaufmann, San Mateo, CA.

Rust, R.T., and R.L. Oliver, 1994, "Service Quality: Insights and Managerial Implications from the Frontier," in R.T. Rust and R.L. Oliver (Eds.), Service Quality: New Directions In Theory And Practice, Sage, London.

Rust, R., and Anthony Zahorik, (1993), "Customer Satisfaction, Customer Retention, and Market Share," *Journal of Retailing*, Vol. 69, No. 2, pp. 193-216.

Rust, R.T., P. Danaher, and S. Varki, 2000, "Using service quality data for competitive marketing decisions", *International Journal of Service Industry Management*, Vol. 11, No. 5, pp. 438-469.

Ryle, G., 1949, The Concept of the Mind. London: Hutchinson.

Scheines, Richard, Peter Spirtes, Clark Glymour, and Christopher Meek, 1994. TETRAD II: Tools for Discovery. Lawrence Erlbaum Associates, Hillsdale, NJ.

Schlosser, A. and A. Kanfer, 1999, "Culture Clash in Internet Marketing: Implications for Marketing Practices," In: Michael Shaw, Robert Blanning, Troy J. Strader, and Andrew Whinston (eds.). Handbook on Electronic Commerce, Springer-Verlag, New York, NY, pp. 195-211.

Schwarz, G., 1978, "Estimating the dimension of a model". *Annals of Statistics*, Vol. 6, No. 2, pp. 461-464.

Schwarz, N., and F. Strack, 1986, "How do respondents answer satisfaction questions? A judgment model of subjective well-being". Paper presented at the annual meeting of the American Association for Public Opinion Research. St. Petersburg, Florida.

Sekaran U., 1992, Research Methods for Business. New York: Wiley and Sons.

Settimi, R., and J. Smith, 1998, "On the Geometry of Bayesian graphical models with hidden variables", In: *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers.

Shankar, V., A. K. Smith, and A. Rangaswamy, 2000, "Customer Satisfaction and Loyalty in Online and Offline Environments", eBusiness Research Center Working Paper 02-2000, October, Penn State University.

Shenoy, C., and P. Shenoy, 1999, "Bayesian Network Models of Portfolio Risk and Return," *Computational Finance*, pp. 87-106.

Shively, Thomas S., Greg M. Allenby, Robert Kohn, 2000, "A Nonparametric Approach to Identifying Latent Relationships in Hierarchical Models." *Marketing Science*, Vol. 19, No. 2, pp. 149-162.

Schumacher, R. E., and R. G. Lomax. 1996. A beginner's guide to SEM. Lawrence Erlbaum Associates, Manwah, New Jersey, USA.

Simon, H., 1994. "Marketing science's pilgrimage to the ivory tower". In: Laurent, G., Lilien, G.L., Pras, B. (Eds.), Research Traditions in Marketing. Kluwer Academic Publishers, Boston, MA, pp. 27–43.

Simon, S. J., 2001, "The impact of Culture and Gender on Web Sites: An Empirical Study", *The DATA BASE for Advances in Inormation Systems*, Vol. 32, No. 1 (Winter), pp. 18-37.

Singh, J. (1991) "Redundancy in Constructs: Problem, Assessment, and Illustrative Example." *Journal of Business Research*, Vol. 22, pp. 255-280.

Singh, J., and D. Sirdeshmukh, 2000, "Agency and Trust Mechanisms in Consumer Satisfaction and Loyalty Judgments," *Journal of the Academy of Marketing Science*, Vol. 28 (Winter), pp. 150-167.

Sirgy, M.J. and Samli, A.C., 1989, "A path analytic model of store loyalty involving self-concept, store image, geographic loyalty, and socioeconomic status", *Journal of the Academy of Marketing Science*, Vol. 13, No. 3, pp. 265-91.

Sitkin, S. B., and N. L. Roth, 1993, "Explaining the limited effectiveness of legalistic 'remedies' for trust/distrust." *Organization Science*. Vol. 4, pp. 367-392.

Smith, B., and T. Rutigliano, 2003. Discover Your Sales Strengths: How the World's Greatest Salespeople Develop Winning Careers. Warner Books.

Song, J. H., G. M. Zinkhan, 2003, "Features Of Web Site Design, Perceptions of Web Site Quality, and Patronage Behavior", ACME 2003 Proceedings.

Sönke, A., 2000, "Impact of functional relationships, decisions, and solutions on the applicability of marketing models." *International Journal of Research in Marketing*, Vol. 17, pp. 169-175.

Spiegelhalter, D. J., and S. L. Lauritzen. 1990, "Sequential updating of conditional probabilities on directed graphical structures". *Networks*, Vol. 20, No. 5 (Special Issue on Influence Diagrams), pp. 579-605.

Spiegelhalter, D.J., A.P. David, S.L. Lauritzen, R.G. Cowell, 1993, "Bayesian Analysis in Expert Systems", *Statistical Science*, Vol. 8, No.3, pp. 219-283.

Spirtes, P., and C. Glymour, 1991. "Algoritm for fast recovery of sparse causal graphs". *Social Science Computer Review*, Vol. 9, pp. 62-72.

Spirtes, P., Glymour, C. and Scheines, R., 2001, Causation, Prediction and Search. Cambridge, MIT Press.

Spreng, Richard A., Scott B. MacKenzie and Richard W. Olshavsky, 1996, "A Reexamination of the Determinants of Consumer Satisfaction," *Journal of Marketing*, Vol. 60, No. 3, July, pp. 15-32.

Srinivas, S., 1993, "A Generalisation of the Noisy-OR Model", In: *Proceedings of the Conference on Uncertainty in AI*, UAI'93, pp. 208-218.

Srinivasan, S.S., R. Anderson and K. Ponnavolu, 2002, "Customer Loyalty in E-commerce: an Exploration of its Antecedents and Consequences", *Journal of Retailing*, Vol. 78, pp. 41-50.

Staw, B.M. 1977, "Two sides of commitment: The rational and irrational components of the Commitment process". Paper presented at the annual meeting of the Academy of Management, Orlando.

Steenkamp, J.-B, 2000, Introduction to the Special Issue on Marketing Modeling on the Threshold of the 21st Century. *International Journal of Research in Marketing*, Vol. 17, No. 2-3, pp. 99-104.

Steenkamp, J.-B., and H. Baumgartner, 2000, "The Use of Structural Equation Models for Marketing Modeling," *International Journal of Research in Marketing*, Vol. 17, No. 2-3, pp. 195-202.

Steiger, J. 1990. "Some additional thoughts on components, factors, and factor indeterminacy." *Multivariate Behavioral Research*, Vol. 25, pp. 41-45.

Stelzl, I., 1986. "Changing the causal hypothesis without changing the fit: Some rules for generating equivalent path models". *Multivariate Behavioral Research*, Vol. 21, pp. 309-331.

Stone, M., 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions". *Journal of the Royal Statistical Society*, Vol. 36, pp. 111-147.

Straub, D. W., 1989, "Validating Instruments in MIS Research," *MIS Quarterly*, Vol. 13, No. 2, pp. 147-169.

Sullivan, J., Peterson, R. B., Kameda, N., & Shimada, J. 1981. "The Relationship Between Conflict Resolution Approaches and Trust – A Cross Cultural Study". *Academy of Management Journal*, Vol. 24, No. 4, pp. 803-815.

Swaminathan, V., Lepkowska-White, E. and Rao, B. P., 1999, "Browsers or Buyers in Cyberspace? An Investigation of Factors Influencing Electronic Exchange," *Journal of Computer-Mediated Communication*, Vol. 5, No. 2.

Szymanski, D. M., and David H. Henard, 2001, "Customer Satisfaction: A Meta-Analysis of the Empirical Evidence", *Journal of the Academy of Marketing Science*, 29 (1), 16-35.

Szymanski, D. M., and Richard T. Hise, 2000, "e-Satisfaction: an Initial Examination", *Journal of Retailing*, Vol. 76, No. 3, pp. 309-322.

Takikawa, M., and B. D'Ambrosio, 1999, "Multiplicative Factorization of Noisy-Max", In: *Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence* (UAI-99), pp. 622-630.

Taylor, Steven A., 1995, "Reconciling Service Quality Measurement with Attitudes," In: Teresa A. Swartz, David E. Bowen, and Stephen W. Brown, (Eds.), *Advances in Services Marketing Management: Research and Practice*, Vol. 4., pp. 1-36. CT: JAI Press.

Teichert, T., and K. Rost, 2003, "Trust, involvement profile and customer retention - modelling, effects and implications". *International Journal of Technology Management*, Vol. 26, No. 5/6.

The International Foundation for Customer Focus, 2000, *The International Foundation for Customer Focus Web Site* [online]. Available from: http://www.ifcf.org/ [last accessed 13 March 2004].

Thiesson, B., 1995, "Accelerated quantification of Bayesian Networks with Incomplete Data", In: U. M. Fayyad, R. Uthurusamy (Eds.), Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal, Canada, August 20-21, 1995. AAAI Press.

Thiesson, B., C. Meek, D. M. Chickering, and D. Heckerman, 1998. "Learning mixtures of DAG models". In: Cooper, G. F. and Moral, S., (eds.), *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pp. 504-513. Morgan Kaufmann Publishers, August 1998.

Thiesson, B., C. Meek, D. M. Chickering, and D. Heckerman, 1999. "Computationally efficient methods for selecting among mixtures of graphical models". In: Bernardo, J., Berger, J., Dawid, A., and Smith, A., (eds.), *Bayesian Statistics*, Vol. 6, pp. 631-656. Oxford University Press.

Thompson, R., D. W. Barclay, and C. A. Higgins, 1995, "The Partial Least Squares Approach to Causal Modeling: Personal Computer Adoption and Use as an Illustration," *Technology Studies*, Special Issue on Research Methodology, Vol. 2, No. 2 (Fall), pp. 284-324.

Tse, D. K., and P. C. Wilton. 1988. "Models of Consumer Satisfaction Formation: An Extension." *Journal of Marketing Research*, Vol. 25 (May), pp. 204-212.

van Bruggen, Gerrit H., Berend Wierenga, 2000. "Broadening the perspective on marketing decision models". *International Journal of Research in Marketing*. Vol. 17, No. 2, pp. 159 – 168.

Van Riel, A. C. R., V. Liljander, and P. Jurriëns, 2001, "Exploring consumer evaluations of e-services: a portal site". *International Journal of Service Industry Management*, Vol. 12, No. 4, pp. 359-377.

Van Riel, A. C. R., J. Semeijn, and W. Janssen, 2003, "E-service quality expectations: a case study", *Total Quaality Management*, Vol. 14, No. 4, pp. 437-450.

Vanhoof, K., and G. Swinnen, 1996, "Attribute Importance: Assessing Nonlinear Patterns of Factors Contributing to Customer Satisfaction," Esomar Publication Series, Vol. 204, pp. 171-183.

Verma, T., and J. Pearl. 1991, "Equivalence and Synthesis of Causal Models." UCLA Cognitive Systems Laboratory, Technical Report (R-150), June 1990, in Proceedings, Workshop on Uncertainty in Artificial Intelligence, July 27-29, 1990. Also in Uncertainty in Artificial Intelligence, Vol. 6, Cambridge, MA, Elsevier Science Publishers, pp. 220-227.

Vermunt, J.K., 1996. "Causal log-linear modelling with latent variables and missing data". In: U. Engel and J. Rei-necke (eds.), Analysis of change: advanced techniques in panel data analysis, pp. 35-60. Berlin/New York: Walter de Gruyter.

Vermunt, J. K., and Magidson, J., 2000. Latent GOLD 2.0 user's guide. [Software manual]. Belmont, MA: Statistical Innovations.

Vermunt, J.K. and Magidson, J., 2004, "Factor Analysis with Categorical Indicators: A Comparison Between Traditional and Latent Class Approaches". In: A. Van der Ark, M.A. Croon and K. Sijtsma (eds), New Developments in Categorical Data Analysis for the Social and Behavioral Sciences. Erlbaum.

Verplanken, B., 1991. "Persuasive communication of risk information: A test of cue versus message processing effects in a field experiment". *Personality and Social Psychology Bulletin*, Vol. 17, pp. 188-193.

Vredenburg, Harrie and Chow-Hou Wee, 1986, "The Role of Customer Service in Determining Customer Satisfaction," *Journal of the Academy of Marketing Science*, Vol. 14, No. 2, pp. 17-26.

von Mises, R., 1928, Probability, Statistics and Truth. Allen & Unwin, London, 1957 (originally published in 1928).

Wedel, M., Kamakura, W. A., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S., & Poulsen, C. S., 1999. "Discrete and continuous representation of heterogeneity." *Marketing Letters*, Vol. 10, pp. 217– 230.

Westbrook, R. A., 1981, "Sources of Consumer Satisfaction with Retail Outlets," *Journal of Retailing*, Vol. 57, No. 3, pp. 68-85.

Westbrook, R. A., 1987, "Product/Consumption Based Affective Responses and Postpurchase Processes," *Journal of Marketing Research*, Vol. 24 (August), pp. 258-70.

Wierenga, B., P.A.M. Oude Ophuis, K.R.E. Huizingh, and P.A.F.M. van Campen, 1994. "Hierarchical scaling of Marketing Decision Support Systems". *Decision Support Systems*. Vol. 12, No. 3, pp. 219-232.

Wierenga, Berend, and Gerrit van Bruggen, 2000. Marketing Management Support Systems: Principles, Tools, and Implementation. Kluwer Academic Publishers.

Woodruff, Robert B., Ernest R. Cadotte and Roger L. Jenkins, 1983, "Modeling Consumer Satisfaction Processes Using Experience-Based Norms," *Journal of Marketing Research*, Vol. 20 (August), pp. 296-304.

Yi, Y., 1990, "A critical review of customer satisfaction," In: V.A. Zeithaml (Ed.), *Review of Marketing*, pp. 68-123. Duke University, AMA.

Zaltman, G., & Wallendorf, M., 1983. Consumer behavior: Basic findings and management implications. New York: John Wiley and Sons.

Zeithaml, Valerie. 1987. "Defining and Relating Price, Perceived Quality, and Perceived Value". Cambridge, Massachusetts: Marketing Science Institute.

Zeithaml, V. A. 1988. "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence." *Journal of Marketing*, Vol. 52 (July), pp. 2-22.

Zeithaml, V.A., Leonard L. Berry, and A. Parasuraman, 1996, "The Behavioural Consequences of Service Quality," *Journal of Marketing*, Vol. 60 (April), pp. 31-46.

Zeithaml, V. A., A. Parasuraman, and Arvind Malhotra. 2000. "e-service Quality: Definition, Dimensions and Conceptual Model." Working Paper. Marketing Science Institute, Cambridge, MA.

Zeithaml, V. A., A. Parasuraman, and Arvind Malhotra. 2002. "Service Quality Delivery Through Web Sites: A Critical Review of Extant Knowledge". *Journal of the Academy of Marketing Science*. Vol. 30, No. 4, pp. 362-375.

Zeithaml, Valarie, 2001. "Service Quality, Profitability, and the Economic Worth of Customers: What We Know and What We Need to Learn", *Journal of the Academy of Marketing Science*, Vol. 28, No. 1, pp. 67-85.

Zellweger, P., 1997, "Web-based Sales: Defining the Cognitive Buyer", *Electronic Markets*, Vol. 7, No. 6, pp. 10-16.

Zweig, G., and S. Russel, 1999, "Probabilistic Modeling with Bayesian Networks for Automatic Speech Recognition." *Australian Journal of Intelligent Information Processing Systems*, Vol. 5, No. 4, pp. 253-60.

Nederlandse samenvatting

Nog maar zeer recent heeft Bayesiaanse data analyse aan interesse gewonnen, onder andere binnen het marketing onderzoeksdomein [Rossi en Allenby, 2003; Wedel *et al.*, 1999; Shively *et al.*, 2000]. Eén van de technieken die gebaseerd is op Bayesiaanse data analyse is Bayesiaanse netwerken. Mathematisch gezien bieden Bayesiaanse netwerken (BN's) een netwerkgebaseerd kader aan dat de gebruiker in staat stelt om modellen die onzekerheid bevatten voor te representeren en te analyseren. Ook is het met Bayesiaanse netwerken mogelijk om tot een exacte en effectieve voorstelling te komen van de gezamenlijke kansverdeling van willekeurige variabelen binnen een bepaald domein. Het Bayesiaans netwerk formalisme is reeds lange tijd ontwikkeld en bekend binnen statistiek. Echter omwille van ernstige rekenkundige problemen geraakte de toepassing in onbruik. Het is dus pas vrij recent dat, nadat de techniek dankzij het werk van Pearl [1988] binnen het domein van artificiële intelligentie opnieuw aan bekendheid inwon, nieuwe succesvolle ontwikkelingen werden voorgesteld. Deze maakten het mogelijk de techniek van Bayesiaanse netwerken toe te passen binnen een breed gamma van domeinen die met onzekerheid te maken hebben [e.g., Andreassen *et al.*, 1991; Heckerman *et al.*, 1992; Heckerman et al., 1995; Jensen et al., 2001; Zweig en Russel, 1999].

De unieke bijdrage van dit proefschrift komt in belangrijke mate voort uit de samenvoeging van de literatuur omtrent Bayesiaanse netwerken met de literatuur omtrent marketing modellering. Ondanks de veelbelovende eigenschappen van Bayesiaanse netwerken om diverse marketingproblemen op te lossen, betreft het nog altijd een eerder onbekende techniek binnen het marketing onderzoeksdomein [Lilien en Rangaswamy, 2000]. Dit gebrek aan erkenning kan in de eerste plaats worden toegeschreven aan het feit dat de methodologie nog steeds in haar beginstadium staat.

Sinds de techniek vanaf 1990 aan populariteit won, heeft het grootste deel van het onderzoek met betrekking tot Bayesiaanse netwerken zich voornamelijk geconcentreerd op de ontwikkeling van algoritmes en het oplossen van problemen binnen het gebied van expertsystemen en datamining. Tot op heden werd er weinig aandacht geschonken aan het toepassen of evalueren van Bayesiaanse netwerken als een potentiële techniek om onderzoek uit te voeren, laat staan marketingonderzoek. Er kan dan ook gesteld worden dat een grondige discussie van de basiskenmerken en van de potentiële toegevoegde waarde van de Bayesiaanse netwerktechnologie in het arsenaal van de marketingonderzoeker tot op heden ontbreekt. De introductie en de motivatie voor dit onderzoek worden daarom besproken in hoofdstuk 1.

Daarom is het algemene doel van dit proefschrift een kritische evaluatie te maken van de toepassing van Bayesiaanse netwerken in theoretisch en praktisch marketingonderzoek. Verder wordt getracht nieuwe methoden en ontwikkelingen binnen Bayesiaanse netwerkmodellering voor te stellen die de huidige mogelijkheden van de techniek, met betrekking tot de specifieke eisen binnen het marketing onderzoeksdomein, verbeteren. Dit werk richt zich verder enkel op één bepaald deel van de marketingwetenschap: namelijk het Klanten Tevredenheids en Getrouwheids-onderzoek (KT&G). Omwille van het stijgende belang van e-commerce en internet binnen marketing [bvb., Mahajan en Venkatesh, 2000; O'Connor en Galvin, 2001], wordt het KT&G fenomeen zowel binnen de traditionele "mortar en brick" als binnen de online context in beschouwing genomen.

De kritische evaluatie die in deze thesis wordt beschreven is eerder intern dan extern gericht. Het doel is om Bayesiaanse netwerken eerder individueel te bestuderen dan in een concurrentieel kader. Het is niet de bedoeling om deze methodologie te vergelijken met andere technieken, die vandaag de dag worden toegepast binnen KT&G onderzoeken, in termen van hun respectievelijke resultaten en bevindingen om alzo na te gaan welke technieken het best zijn. Bijgevolg luidt de stelling, die in deze scriptie wordt aangenomen, dat de Bayesiaanse netwerkbenadering eerder een andere aanpak is die kan bijdragen om het KT&G fenomeen beter te begrijpen.

Het verschil tussen het theoretisch en praktisch onderzoek is afgebakend volgens de positieve/normatieve dimensie binnen de marketingwetenschapsfilosofie. De focus van het theoretische KT&G onderzoek wordt gedefinieerd als de identificatie van cognitieve, affectieve en normatieve processen na aankoop waardoor klanten tevreden worden en eventueel ook trouw gaan zijn aan een diensten- of productaanbieder. Anders gezegd is het een wetenschappelijk onderzoek dat tot doel heeft om de theorie van het klantentevredenheids- en getrouwheidsfenomeen te ontwikkelen. Het doel van het praktisch KT&G onderzoek is om een aantal specifieke problemen, waar een bedrijf mee te maken kan hebben, op te lossen. Kenmerkend is dat de analyse zich concentreert op de relatie tussen het belang van dienst- en productattributen om de algemene klantentevredenheid met de dienst of het product te verklaren. Deze relatie kan variëren tussen bedrijven en is sterk afhankelijk van de unieke karakteristieken van de dienst en/of het product en van de industrie. Deze analyse wordt dikwijls gedefinieerd in de marketingliteratuur als de belang/prestatie analyse [bvb. Martilla, 1977] en er wordt in deze scriptie naar verwezen als praktisch onderzoek naar klantentevredenheid.

Vervolgens worden de specifieke vereisten, met betrekking tot statistische technieken die worden toegepast in theoretisch en/of praktisch KT&G onderzoek, voorgesteld. Op basis van deze vereisten, beschrijft de auteur het algemene doel van de thesis in termen van meer gedetailleerde deeldoelstellingen. Deze deeldoelstellingen worden afzonderlijk gedefinieerd voor het gebruik van BN in theoretisch en praktisch onderzoek. De deeldoelstellingen met betrekking tot het gebruik van BN in theoretisch KT&G onderzoek worden voorgesteld in Tabel 1.

	Case	
	1	2
1. Hoe kunnen marketing theorieën worden ontwikkeld met behulp van de Bayesiaanse netwerkbenadering?		
a. Studie van Bayesiaanse netwerken in verschillende scenario's:		
i. inductief	√	
ii. deductief		√
b. Voorstelling en evaluatie van nieuwe methoden om te gaan met structurele en meetmodellen, die in het bijzonder tot doelstelling hebben om:		
i. Het meetmodel verklaren		√
ii. Latente constructs valideren		√
iii. De beste dimensionering van latente constructs te vinden		√
c. Studie en bespreking van specifieke vraagstukken in de ontwikkeling van theorieën:		
i. Het modelleren van moderator effecten	√	
ii. Rekening houden met bemiddelende variabelen.	√	
2. In welke mate zijn marketingtheorieën ontwikkelt met Bayesiaanse netwerken, onderhevig aan wetenschappelijke validatie? Hoe kunnen ze wetenschappelijk worden gerechtvaardigd (gevalideerd)?		
a. Evalueren van het beschrijvend, voorspellend en verklarend potentieel van Bayesiaanse netwerken in het voorbeeld van het e-tevredenheids en getrouwheidsdomein	√	
3. Wat is de toegevoegde waarde van marketing problemen te modelleren met behulp van Bayesiaanse netwerken?		
a. Aantonen van de mogelijkheden aan van probabilistisch redeneren (voorwaarts, achterwaarts, intercausaal) in het domein	√	
b. Aantonen van het potentieel aan van het uitvoeren van "what-if" simulaties	√	
c. Aantonen van het potentieel aan van de combinatie van voorafgaande kennis met data	√	√
4. Nagaan van de sterkten en zwakten van Bayesiaanse netwerken in termen van specifieke technische en statische modelleringonderwerpen, zoals verdelingskarakteristieken, het omgaan met ontbrekende waarden, enz	√	√

Tabel 1. Deel-doelstellingen in het deel van theoretisch KT onderzoek.

Met betrekking tot praktisch klantentevredenheidsonderzoek werd één onderzoeksvraag voorgelegd dat tot algemeen doel heeft om Bayesiaanse netwerken in deze stroming van marketingonderzoek te evalueren. De deeldoelstellingen binnen deze vraag worden voorgesteld in Tabel 2.

De onderzoeksstrategie wordt besproken in hoofdstuk 1 en bestaat uit 4 gevalstudies, waarvan twee in elk deel. Elke gevalstudie wordt gedefinieerd in termen van de verschillende onderzoeksvragen en deel-doelstellingen zoals ze hierboven werden weergegeven. Om deze doelstellingen te verwezenlijken heeft elke gevalstudie nood aan empirische data. De data die worden gebruikt zijn bestaande, secundaire data die worden verzameld door marktonderzoeksbureaus in België en in Nederland ten dienste van KT&G metingprogramma's voor hun klanten.

	Case 3	Case 4
1. Hoe kunnen Bayesiaanse netwerken worden toegepast in diensten karakteristieken/dimensie belang/prestatie studie?		
a. Pas Bayesiaanse netwerken aan en onderzoek de toepassing ervan in de classificatie van dienstendimensie analyse, met als voornaamste doelen:		
i. De identificatie van het afgeleide belang van dienstendimensies voor de algemene (on)tevredenheids beoordeling	√	
ii. Ondersteunen van marketingbeslissingen door middel van belang/prestatie analyse	√	
iii. Ontdekken van interactie-effecten (synergie en negatie) tussen dienstendimensies	√	
b. Pas Bayesiaanse netwerken aan en onderzoek de toepassing ervan in de classificatie van dienstenkarakteristieken (variabelen):		
i. Evalueren van een model met mediators van algemene tevredenheid, gebaseerd op de techniek van "parent divorcing" in de analyse van de belangrijkheid van karakteristieken		√
ii. Nagaan of het mogelijk is om, in een mediërend model, tevredenheid met de dienstendimensie te behandelen als een verborgen node, en dus een vragenlijst te optimaliseren door geen vragen te stellen omtrent tevredenheid binnen de dienstendimensie		√
iii. Evalueren van het "noisy-OR" model van algemene tevredenheid in de analyse van de belangrijkheid van karakteristieken		√
2. Nagaan van de sterkten en de zwakten van Bayesiaanse netwerken in termen van specifieke technische en statistische modelleringonderwerpen, zoals veronderstellingen over de dataverdeling, het omgaan met ontbrekende waarden, enz	√	

Tabel 2. Deel-doelstellingen in het deel van praktisch KT onderzoek.

In hoofdstuk 2 volgt een bespreking voor van de belangrijkste onderwerpen binnen het gebruik van Bayesiaanse netwerken op basis van bestaande machineleren en datamining literatuur. Meer specifiek wordt de historische achtergrond van het gebruik van Bayesiaanse netwerken binnen onderzoeksdomein van expertsystemen besproken. Vervolgens wordt een korte introductie tot kansberekening en grafische theorie gegeven en worden Bayesiaanse netwerken formeel gedefinieerd. Verschillende methodes om Bayesiaanse netwerken te bouwen, waaronder het bouwen van netwerken op basis van voorafgaande kennis en het leren van netweken uit data, worden voorgesteld. De bespreking behandelt vooral deze methoden uit de BN literatuur dewelke het meest relevant zijn voor het vervolg van het proefschrift. Vervolgens worden de meest belangrijke onderwerpen om de kwaliteit van Bayesiaanse netwerkmodellen te beoordelen besproken. Er wordt afgesloten met een bespreking van hun praktisch gebruik, gevolgd door de conclusies.

Zowel conceptuele definities als theoretische antecedenten en consequenten van theoretische constructies, die aanwezig zijn in de gevalstudies, worden besproken in het eerste deel van hoofdstuk 3, dat gebaseerd is op de KT&G literatuur. Het tweede deel van dit hoofdstuk bevat een korte bespreking van de gebruikte methoden in theoretisch en praktisch KT&G onderzoek, i.e. klassieke regressie, SEM en PLS modellen. Aandacht wordt gegeven aan hun statistische onderbouw, specificatie, schatting, evaluatie, interpretatie en hun beperkingen. Dit hoofdstuk sluit af met conclusies in functie van elk van de gevalstudies die zullen worden besproken.

Het eerste probleem dat wordt aangepakt in de gevalstudie in hoofdstuk 4 is hoe marketingtheorieën, op basis van de inductieve benadering tot onderzoek, kunnen worden ontdekt door middel van Bayesiaanse netwerken. De tweede vraag die hier wordt behandeld is hoe marketingtheorieën, die worden ontdekt met BNs, wetenschappelijk kunnen worden gerechtvaardigd of gevalideerd. Concreet betekent dit dat wordt nagegaan of de uitkomst van dit inductieve proces, i.e. het voorgestelde theoretisch model van e-loyalty, aan de criteria voldoet om als wetenschappelijk gefundeerde theorie beschouwd te worden. Om hierop te kunnen antwoorden wordt gebruik gemaakt van de benadering van Hunt [1991]. Om de verklarende sterkte van een Bayesiaans e-loyalty model op een meer systematische manier te onderzoeken, zal het veronderstelde verklarend potentieel van dit model worden onderzocht. Dit onderzoek zal, zoals aanbevolen door de moderne empirische benadering in de wetenschapsfilosofie [Hunt, 1991], gebeuren door de criteria, dewelke werden ontdekt, te evalueren:

- er moet worden aangetoond dat het fenomeen, dat verklaard dient te worden, enigszins verwacht werd te gebeuren;
- dat het model intersubjectief verifieerbaar is;
- dat het model een empirische inhoud heeft;
- dat het model pragmatisch is.

Verder worden eveneens belangrijke vereisten van technieken, die tot doel hebben om bij te dragen tot de wetenschappelijke begrip van marketing fenomenen en e-loyalty in het bijzonder, zoals de onderwerpen van moderator- en mediator variabelen [Bagozzi, 1994a] besproken en geëvalueerd. De volgende vraag die in dit hoofdstuk gesteld wordt luidt: "wat is de toegevoegde waarde van het modelleren van marketingproblemen met behulp van Bayesiaanse netwerken?"

In hoofdstuk 5, wordt een gevalstudie waarin de BN benadering geëvalueerd binnen het deductief KT&G onderzoek. Onze veronderstelling betreffende het KT&G fenomeen kan worden beschouwd als een deel van het literatuuroverzicht, dat reeds werd voorgesteld in hoofdstuk 3. Daarom wordt dit hoofdstuk toegespitst op de overblijvende stappen in het proces. Er worden mogelijke hypotheses voorgesteld betreffende de af- of aanwezigheid van directe relaties tussen concepten. Ook worden modellen gevormd en met elkaar vergeleken door middel van de posterior kans-maatstaf. Ten tweede worden nieuwe methoden voorgesteld en geëvalueerd die de Bayesiaanse netwerkbenadering aanvullen met de mogelijkheid om rekening te houden met latente construct en metingsmodellen. Als een eerste onderwerp in dit opzicht, wordt een specifieke methode voorgesteld om met latente constructs en met een structureel model om te gaan, alsook om rekening te houden met het meetmodel in Bayesiaanse netwerkmodellering. Meer specifiek bestaat het idee erin om latente constructs expliciet te integreren in het meetmodel door het gebruik van een speciaal type van Bayesiaans netwerkmodel, i.e. Naïve Bayes structuren [Duda en Hurt, 1973]. Als volgend onderwerp wordt een methode voorgesteld om latente constructs te valideren binnen de Bayesiaanse netwerktechnologie. De constructvalidatie benadering, dewelke in deze studie wordt voorgesteld, kan worden gezien als de mate waarin een operationalisering in staat is om het concept te meten dat het wordt geacht te meten [bvb, Cook en Campbel, 1979]. In onze implementatie wordt geëvalueerd of de indicatorvariabelen eerder aan één potentiële construct of aan meerdere verschillende potentiële constructs gerelateerd zijn. Bovendien bestaat een andere deeldoelstelling van deze studie in het voorstellen en evalueren van een methode die de cardinaliteit kan bepalen van latente constructs in Bayesiaanse netwerkmodellen. Vervolgens wordt door verschillende prior verdelingen van conditionele kansen toe te laten, het potentieel aangetoond van de combinatie van voorafgaande kennis met beschikbare data. Tot slot wordt, bij de bespreking, de sterkten en zwakten van Bayesiaanse netwerken, in termen van specifieke statistische en modellerings-onderwerpen zoals veronderstellingen omtrent gegevens verdelingen, omgaan met ontbrekende waarden, etc., aangetoond.

De gevalstudie in hoofdstuk 6 handelt over het gebruik van BNs in praktisch KT onderzoek. Eerst worden Bayesiaanse netwerken aangepast en bestudeerd met als doel om het afgeleid belang van de potentiële factoren voor de algemene (on)tevredenheidsbeoordeling te identificeren. De doelstelling is om na te gaan welke diensten/product dimensies potentiële bronnen van (on)tevredenheid kunnen zijn. Om dit te bereiken wordt een procedure toegepast dewelke op sensitiviteitsanalyse in Bayesiaanse netwerken gebaseerd is. Ten tweede wordt, door middel van belangrijkheid-prestatie analyse, de Bayesiaanse netwerkbenadering geëvalueerd met het oog op marketingbeslissingen. Het doel van deze analyse is om aan te tonen op welke dienstendimensies een bedrijf zich in de eerste plaats zou moeten richten, en welke dimensies onderwerp zijn van mogelijke oververzorging. Een aantal van de categorieën die worden gedefinieerd zijn: lage prioriteit, optreden vereist, kansen, sterkten, zorg dragen en mogelijke oververzorging. Het derde onderwerp dat wordt beschouwd is of en in welke mate Bayesiaanse netwerken kunnen worden toegepast om interactie-effecten tussen dienstendimensies te ontdekken. Tot slot, maar daarom niet minder belangrijk, worden, in termen van statistische en modelleringsonderwerpen, de sterkten en zwakten van Bayesiaanse netwerken onderzocht, e.g. door middel van een optimaal gebruik van alle beschikbare data in één model.

In hoofdstuk 7 wordt een andere gevalstudie in praktische KT studies onderzocht. Ten eerste, wordt een model met mediator variabelen van algemene tevredenheid geëvalueerd op basis van de techniek van "parent divorcing". Meer precies bevat het model een klantentevredenheidsbeoordeling, dewelke geoperationaliseerd wordt door een klantenvragenlijst, alle variabelen zoals ze worden geobserveerd. De evaluatie van dit model zal gebeuren op basis van twee criteria, gebaseerd op 1) de mogelijkheid om een belangrijke classificatie van de dienstenkenmerken door te voeren, en 2) de mogelijkheid om de algemene mate van tevredenheid te voorspellen. Als een alternatief voor het model met dienstendimensies, waarbij de waarden onmiddellijk worden geoperationaliseerd, wordt een model beschouwd waarbij alle dienstendimensies als verborgen nodes worden beschouwd. In dit verborgen construct model, kunnen al de nodige kansen worden geschat op basis van de overblijvende variabelen en de afhankelijkheden die door het model worden verondersteld door middel van een optimalisatietechniek, zoals bvb het EM algoritme. Daarom betreft een tweede belangrijke onderzoeksvraag, die in deze studie wordt beschouwd, of het noodzakelijk is om klantentevredenheid, met een product en/of dienst, te meten door dit rechtstreeks te bevragen met behulp van een vragenlijst. Misschien is het mogelijk om de vragenlijst te optimaliseren zonder een bevraging te doen naar de tevredenheid binnen de dienstendimensie. Uit de scores, betreffende de algemene tevredenheid en tevredenheid met specifieke kenmerken, moet dan kunnen worden afgeleid wat de invloed is van de dienstendimensies.Om dit te beoordelen, werden een aantal alternatieve modellen van klantentevredenheid vergeleken: 1) één waarbij het tevredenheidniveau met een dienstendimensie indirect wordt afgeleid uit de data door middel van een maximum likelihood schatting en 2) één waarbij dit niveau expliciet wordt gemeten en er mee rekening wordt gehouden door middel van de vragenlijst. Om te bepalen of de twee modellen equivalent zijn in praktische tevredenheidstudies, zullen 2 types van vergelijkende validatie gebruikt worden. Ten eerste zullen voor kwalitatieve validatie de resultaten van classificatie, van de kenmerken die gebruik maken van het, in de vorige gevalstudie ontwikkelde, raamwerk, worden vergeleken. Ten tweede zal de voorspelde accuraatheid dienen als een tweede type van validatie. Verder is deze gevalstudie er op gericht om andere afhankelijkheden van verdelingen te beoordelen, zoals ondermeer het "noisy OR-gate" voor de analyse van de belangrijkheid van de karakteristieken.

Besluiten, implicaties en beperkingen van dit proefschrift worden besproken in hoofdstuk 8. Met betrekking tot het gebruik van Bayesiaanse netwerken in theoretisch KT&G onderzoek, kunnen we besluiten dat de Bayesiaanse netwerkbenadering theoretisch goede resultaten geeft voor causale gevolgtrekkingen uit data. Immers, de resultaten van het theoretisch model op basis van de 4 verschillende datasets brengen vergelijkbare besluiten met zich mee, dewelke het bestaan van een algemeen theoretisch model van e-loyalty lijken te suggereren. Alsook de resultaten van de tweede gevalstudie in klantengetrouwheid bevestigen het *a priori* voorgestelde theoretisch model van dit verschijnsel. Dit suggereert dat de modelvalidatie procedure, dewelke gebaseerd is op de posterior kans van een model, een waardevolle manier is om zowel de theorie van klantentevredenheid en getrouwheid te ontdekken en te bevestigen. Bovendien zijn zowel de inductieve als de deductieve benadering geschikt voor gebruik met Bayesiaanse netwerken.

Verder wordt geconcludeerd dat de Bayesiaanse netwerkbenadering beschouwd kan worden als een techniek dewelke een wetenschappelijk gefundeerde theorie oplevert. Deze methodologie werd als geschikt bevonden en kan worden gebruikt om klantentevredenheid- en getrouwheidstheorieën te modelleren en om wetenschappelijke begrijpbaarheid van dit verschijnsel op een empirische basis af te leveren. Er wordt ook gepleit dat de ware toegevoegde waarde van Bayesiaanse netwerken bestaan uit de unieke kenmerken van kansredenering en "what-if" simulaties, alsook het potentieel om bestaande kennis omtrent het KT&G fenomeen te combineren met data, om alzo te komen tot verbeterde theorie ontdekking en validatie.

Er kan geconcludeerd worden dat de techniek, dewelke wordt voorgesteld op basis van sensitiviteitsanalyse in Bayesiaanse netwerken, gebruikt kan worden in een karakteristieken/dimensies belangrijkheden/prestatie analyse met betrekking tot dienstenkenmerken. Bovendien kan de voorgestelde techniek worden gebruikt om interactie-effecten te ontdekken tussen dienstenkenmerken. Met betrekking tot het onderwerp van vragenlijstoptimalisatie, moet worden besloten dat onze benadering het niet mogelijk maakt om geen bevraging te doen omtrent tevredenheid met dienstendimensies.

Bij de samenvatting van de sterkten en zwakten, wordt beargumenteert dat de Bayesiaanse netwerkbenadering, zoals ze wordt toegepast in de context van KT&G onderzoeken, meer sterktes dan zwaktes aanbiedt. Een greep uit de meest belangrijke sterktes die we kunnen vermelden: het potentieel van een theoretisch adequaat omgaan met ontbrekende waarden, het potentieel om met vele verschillende kansverdelingen van data om te gaan, en het gemak van gebruik en interpretatie van de techniek door niet-experten. De belangrijkste zwaktes zijn volgens ons, de afwezigheid van een volledig uitgewerkte procedure van structurele en metingmodellering; onze benadering maakt het immers niet mogelijk om de metingsfout te controleren. De benadering in dit werkstuk zou dan ook eerder moeten worden beschouwd als een eerste poging om met deze zwakte om te gaan, maar het lost deze zeker niet volledig op.

Dit proefschrift beidt een eerste kritische evaluatie van de BN benadering in de context van KT&G onderzoek. Meer inspanningen moeten worden verricht in de toekomst om de toegevoegde waarde van deze benadering te bevestigen en dit in een concurrentieel kader met andere technieken.