Bayesian Network Analysis of Brand Concept Maps

by

Steven Alan Brownstein

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2013 by the
Graduate Supervisory Committee:

Peter Reingen, Chair
Ajith Kumar
Michael Mokwa

ARIZONA STATE UNIVERSITY

December 2013

ABSTRACT

We apply a Bayesian network-based approach for determining the structure of consumers' brand concept maps, and we further extend this approach in order to provide a precise delineation of the set of cognitive variations of that brand concept map structure which can simultaneously coexist within the data. This methodology can operate with nonlinear as well as linear relationships between the variables, and utilizes simple Likert-style marketing survey data as input. In addition, the method can operate without any *a priori* hypothesized structures or relations among the brand associations in the model.

The resulting brand concept map structures delineate directional (as opposed to simply correlational) relations among the brand associations, and differentiates between the predictive and the diagnostic directions within each link. Further, we determine a Bayesian network-based link strength measure, and apply it to a comparison of the strengths of the connections between different semantic categories of brand association descriptors, as well as between different strategically important drivers of brand differentiation. Finally, we apply a precise form of information propagation through the predictive and diagnostic links within the network in order to evaluate the effect of introducing new information to the brand concept network.

This overall methodology operates via a factorization of the joint distribution of the brand association variables via conditional independence properties and an application of the causal Markov condition, and as such, it represents an alternative approach to correlation-based structural determination methods. By using conditional independence as a core structural construct, the methods utilized here are especially well-

suited for determining and analyzing asymmetric or directional beliefs about brand or product attributes.

This methodology builds on the pioneering Brand Concept Mapping approach of Roedder John et al. (2006). Similar to that approach, the Bayesian network-based method derives the specific link-by-link structure among a brand's associations, and also allows for a precise quantitative determination of the likely effects that manipulation of specific brand associations will have upon other strategically important associations within that brand image. In addition, the method's precise informational semantics and specific structural measures allow for a greater understanding of the structure of these brand associations.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

1. Directionality and Brand Concept Maps

1.1 Consumer Belief Mapping and the BCM Methodology

Marketing researchers and practitioners have long utilized geometric and graphical representations as a means of understanding consumers' product and brand-related beliefs as well as to segment consumers into psychographic clusters which are related to these differing beliefs (Wells, 1985; Dillon, Madden, and Firtle, 1987). Furthermore, an understanding of the dimensionality and structure of consumers' product and brand images is known to be a critical component of a company's strategic marketing mix, and is at the core of firms' efforts to differentiate their products and brands, as well as to establish sustainable brand equity (Park, Jaworski, and MacInnis, 1986; Park, Milberg, and Lawson, 1991; Keller, 1993; Aaker, 1996). Marketing practitioners have also realized the competitive advantage to be gained by designing and promoting products to fit specific regions within consumers' perceptual space for the relevant product category (Morgan and Purnell, 1969; Klahr, 1970; Shocker and Srinivasan, 1974; Huber and Holbrook, 1979; Shocker and Srinivasan, 1979; Hauser and Simmie, 1981).

Many different multivariate statistical methodologies have been utilized to obtain a graphical or geometric representation of consumers' perceived market, brand, and product structures. For example, Gwin and Gwin (2003) have compiled a comprehensive typology of multivariate techniques which have been applied for this purpose in the marketing domain, including multidimensional scaling, factor analysis, correspondence analysis and optimal rescaling, principal component analysis, and discriminant analysis, among others.

Although these various multivariate methodologies produce informative spatial mappings of consumers' brand and product concepts, such techniques do not necessarily reveal the link-by-link pattern of connectivity among brand associations which presumably give rise to the similarities and differences which are perceived to exist among brand or product attribute or among the brands and products themselves (Roedder John et al., 2006). While it is true that the similarity matrix on which many of these scaling techniques are based does presumably derive from semantic relatedness of the given set of associations, Roedder John et al.'s point is valid, namely that even with a specific set of similarity ratings, one is still unsure as to the specific association patterns that gave rise to these similarities. For instance, if two brand associations have a certain degree of rated similarity, one would not necessarily know whether such similarity comes about through a direct connection between these associations, through a series of intermediary associations involving other domain variables, or via a combination of these various forms of interconnection.

A core principle of brand management is that a brand's main locus of differentiation from its competitors resides within the network of perceived associations to that brand (Keller, 1993). Furthermore, it is the structure of that network of brand associations (and not merely its content alone) which constitutes the brand's image for consumers (Aaker, 1996). Hence, without knowledge of the exact link-by-link structure of the network of brand associations, marketing practitioners are often left without sufficient actionable intelligence to determine which variables to manipulate in order to affect particular strategically relevant variables which comprise their brand's image.

An alternative to such multivariate dimensional analysis and imaging techniques is represented by qualitative techniques such as projective studies (Kassarjian, 1974; Rook, 2006), metaphor elicitation methodologies such as the ZMET (Zaltman and Coulter, 1995; Zaltman, 1997) along with several related qualitative techniques (Levy, 1985; Levy, 2006). As applied within the consumer behavior domain, such approaches typically utilize a combination of in-depth interviews and projective probes in order to identify brand associations, and then these associations are typically pieced together into a network of brand or product-related beliefs through in-depth post-hoc analyses of the recovered associations (e.g., Zaltman and Coulter, 1995). Furthermore, a very common research technique involves using such projective or metaphor-based techniques to elicit a collection of brand or product associations, followed by the application of large-scale survey-based research to quantify the prevalence of each uncovered belief within the target population (Dillon, Madden, and Firtle, 1987).

While it is clear that marketing researchers and practitioners can utilize such qualitative techniques to uncover deeper product and brand meanings and associations, and then follow this up with quantitative (typically survey-based) studies which attempt to generalize these findings by exploring their strength and boundary conditions (Dillon, Madden, and Firtle, 1987), there have also been attempts to more directly embed qualitative techniques within a more general quantitative consumer research framework. One quite notable stream of research in this area is that of hierarchical value mapping (or means-end chains), which seeks to couple quantitative methods with a qualitative link-by-link understanding of consumer values and goals as they relate to product attributes

(Reynolds and Gutman, 1988; Gutman, 1997). However, these methods are specific to determining the structure of links between levels within a needs hierarchy model for specific products or product categories. Since such methods are tied to this specific form of psychological model, it is unknown how they would generalize to the wider setting of brand concepts in general.

Roedder-John et al. (2006) discuss a class of techniques which they term ‘analytical mapping’, as exemplified by the work of Henderson et al. (1998), which utilizes network analysis algorithms to quantitatively derive the link-by-link interdependence structure of brand associations.[1] However, such analytical mapping techniques typically utilize discrete cutoff values to decide when a link exists or does not exist between a brand and an association or between two associations, and hence these methods do not as of yet utilize the power of statistical analysis to aid in determining the robustness or comparative strength of the derived associations.

A critical development within the consumer belief mapping literature has been the work of Roedder John et al. (2006). These researchers have developed a unique methodology which not only uncovers the link-by-link structure of consumers’ brand concept maps, but does so in a way that combines various aspects of qualitative methods such as the ZMET technique with concept mapping methodologies from the social and physical sciences (e.g., Ruiz-Primo and Shavelson, 1996). Furthermore, the method of Roedder John et al. (2006) also provides a determination of comparative strengths of the

[1] Although less well-known, the Galois lattice analysis methodology makes a posit that is similar to that used within network analysis, namely that the interconnections between brand or product associations can arise via mutual embeddedness or instantiation within the same exemplar or product (Brownstein, Sirsi, Ward, and Reingen, 2000).

various inter-association links within the derived brand concept map, as well as a means of aggregating individual maps into a consensus map of the focal domain.

The methodology of Roedder John et al. (2006) first employs open-ended interviews which establish a corpus of elicited brand associations, and a first round of aggregation is incorporated at this stage by selecting only those brand associations that are mentioned by at least 50% of respondents. The next step in this methodology directly probes consumers' beliefs about the connections between brand associations. Subjects at this stage are presented with the aggregated list of elicited brand associations, and are asked to assemble them into a network, using either one, two, or three links between associations in order to express their perceived strength of that connection.[2] Finally, a post-hoc aggregation procedure is applied to the individual concept maps produced by the respondents in order to derive a consensus map of the domain, which the authors term a Brand Concept Map (or BCM).

To aggregate the individual maps into a BCM, this technique utilizes a fairly complex five-stage procedure. To begin with, core brand associations are identified as those which are either included on at least 50% of the respondents' maps, or which were included on 45% to 49% of those individual maps, but which also had an interconnection count that was higher than the other identified core brand associations. Next, the identified core brand associations which have ratios of first-order mentions to total mentions of at least 50% and which concurrently have more superordinate than

[2] Blank cards are also provided at this stage so that individuals may add any additional associations which they feel are necessary, but which are not necessarily on any of the cards in the aggregate list provided. Once again, a 'frequency of mention' cutoff was used on these individually added associations.

subordinate connections are selected to be linked directly to the brand (so-called ‘first-order associations’), and then the remaining core associations are linked to these first-order associations if that particular concept-to-concept link was present in at least five individual concept maps. (This criterion is also used to establish links between the first-order associations as well.) Finally, non-core associations are added in, and are linked to the core associations using a cutoff frequency for the number of times the concept-to-concept link is included on individual maps. The final strength of each concept-to-concept link is established as the average of the number of ‘link bars’ (either one, two, or three) that the subjects had ascribed to each particular link which was included in the final consensus map, or BCM.

The final aggregate Brand Concept Map (BCM) for patients’ perceptions of the Mayo Clinic brand, as derived by Roedder John et al. (2006, Figure 3A), is shown in Figure 1 below. As one can see, each variable is connected via (typically multiple) pathways to every other variable in the network. While this high degree of interconnectivity is capable of portraying the intricate nature of the brand concept network topology, it also poses a serious problem for inference within such a structure, since one seemingly has no principled means of resolving the multiple (possible competing) effects of the different pathways between specific variables. These concerns are further addressed in the following section.

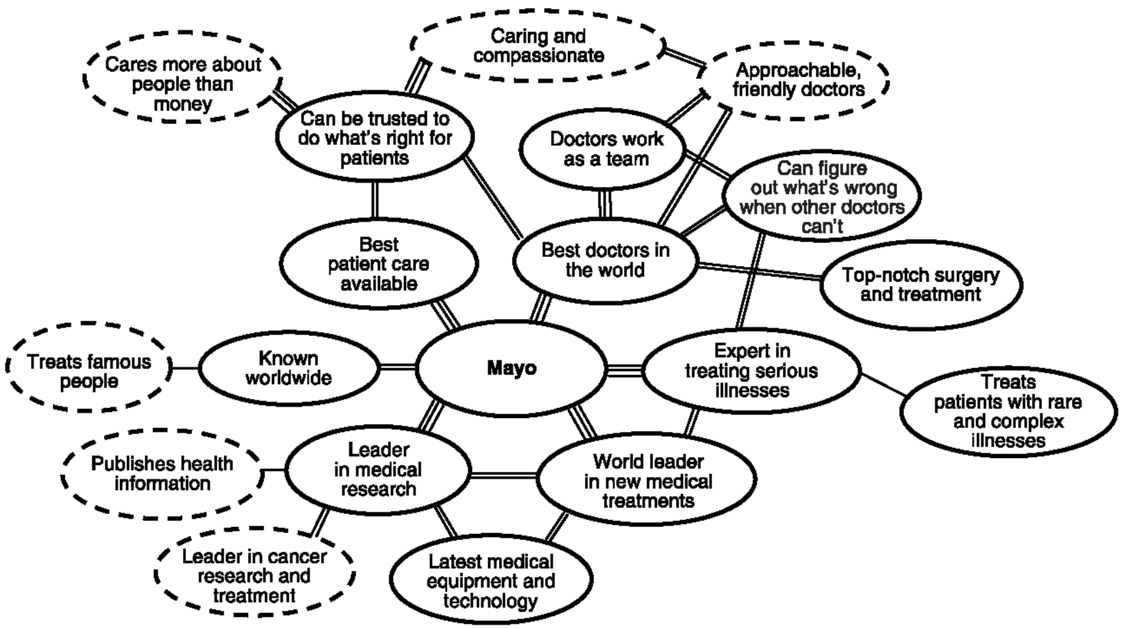


Figure 1. BCM for patients' perceptions of the Mayo Clinic brand.

From Roedder-John et al., 2006, p. 556.

1.2 Extending the BCM

Analytical and cognitive tractability. Consumers' perceptions of products and brands are known to be exceedingly multidimensional and highly intricate constructs (Holbrook and Hirschman, 1982; Alba and Hutchinson, 1987; Glazer and Nakamoto, 1991), and the structures derived via the BCM methodology certainly exhibit a high degree of interconnectedness and complexity. In essence, every variable within such a model is associated with all others, either through a direct link or through a connected series of links, and it is difficult to resolve the relative contributions of all the different pathways of influence between the brand associations in such a model. In turn, inference and evidence propagation are exceedingly difficult to analyze within such structures.

As an example of these multiple pathways of influence among variables, consider that within the BCM for the Mayo Clinic brand (Figure 1), we see that "*Expert in treating serious illnesses*" is linked to "*World leader in new medical treatments*" through both a direct pathway between these two variables, as well as through the link to the core brand "*Mayo*" and then from there to "*World leader in new medical treatments*". Furthermore, the influence of "*Expert in treating serious illnesses*" can also travel 'upwards' in the network (i.e., directed towards the top of the graph) through multiple pathways and then feeding back down through "*Mayo*" and finally into "*World leader in new medical treatments*", or it can travel through the core "Mayo" association and proceed 'downwards' in the graph (e.g., through "*Leader in medical research*") and eventually circle around (through two different pathways) to "*World leader in new medical treatments*".

Analytically, it is difficult to resolve the differing contributions of these multiple pathways of influence. In a parallel fashion, it seems that consumers' ability to cognitively function within such a complex structure would also be very limited. For instance, a consumer routinely encounters tens of thousands of products in just a typical supermarket trip (Food Marketing Institute, 2013). In fact, just the category of toothpaste alone routinely has over one hundred exemplars within a typical American supermarket (Broniarczyk, 2006). If a consumer had to cognitively access the myriad product and brand associations encountered on such a shopping excursion via a fully connected undirected network (in which each product or brand association is connected to all others, either directly or through a chain of associations), such a consumer would have an unimaginably difficult cognitive task to perform in order to make general deductions, inferences, and choices within such an immense network (e.g., Pearl, 1988). Hence one of the central goals of the extension to techniques such as the BCM which is proposed in this thesis will consist of a means to make such structures more computationally and cognitively tractable, without in any way diminishing their applicability or universality.

Brand-specific associations. A second area in which we wish to extend the BCM methodology is in the form of data collection that is utilized and the types of links that are retrieved. Specifically, at the link elicitation stage, respondents in the BCM are asked to arrange a set of index cards containing brand associations into an overall map or network that portrays what they think of the core brand. The respondents are instructed to place a link between two brand characteristics if they feel that the brand itself is characterized by such an association. This may actually be quite a subtle distinction for respondents to make. For instance, if a respondent draws a link between two particular brand attributes, it may be difficult to determine whether that respondent is asserting that the brand is characterized by *that particular association between the two attributes*, or merely that the brand is perceived to *possess* each of these attributes and the attributes are semantically related to one another in a wider context that is not necessarily specific to the brand being examined.

For example, the BCM analysis of the Mayo Clinic brand (from Roedder John, et al., 2006) involves the brand associations that are listed in Table 1. To illustrate the potential methodological difficulty described above, suppose that a respondent in the procedure is examining the particular brand association "*Cares more about people than money*", in order to see where to fit this association into the network that he or she is currently constructing. Such a respondent might end up linking this association to other such associations that deal with either the general concept of 'money', the category of 'people', or the general concept of 'caring', since those are the main semantic categories that are mentioned in the statement. Perhaps such a respondent would select "*Caring and compassionate*", or "*Can be trusted to do what's right for patients*". However, would such a connection between concepts reflect what the respondent truly feels about the Mayo clinic as a brand, or would they merely reflect the fact that the respondent has connected statements that seem semantically related in general?

Clearly, it is quite possible that people may conflate certain aspects of their 'true' opinion of the Mayo clinic as a brand with equally salient (and possibly much more cognitively available) reasoning patterns about the general topic at hand. To the extent that this is the case, portions of the consensus map within a BCM may merely reflect a culturally shared understanding of which brand associations are semantically related to one another in a manner that is not necessarily brand-specific. Since customer-based brand equity is critically dependent upon the set of *unique* associations held by the brand (Aaker, 1991; Keller, 1993), the possible conflation of brand-specific inter-attribute associations with attribute associations that may simply reflect respondents' general culturally-based understanding of the semantic relatedness of the probes themselves can raise concerns about the interpretability of the derived structures.

Table 1

List of Brand Associations Used in the BCM Study of the Mayo Clinic

Expert in treating serious illness
Latest medical equipment and technology
Leader in medical research
Known worldwide
Top-notch surgery and treatment
Best doctors in the world
World leader in new medical treatments
Can be trusted to do what’s right for patients
Doctors work as a team
Best patient care available
Treats patients with rare and complex illnesses
Can figure out what’s wrong when others can’t
Publishes health information to help you stay well
Approachable, friendly doctors
Caring and compassionate
Treats famous people from around the world
It comforts me knowing Mayo exists if I ever need it
People I know recommend Mayo
Leader in cancer research and treatment
Cares more about people than money
Court of last resort
Hard to get into unless very sick or famous
Very big and intimidating
Expensive
Uses its reputation to make money

Note. From Roedder John et al. (2006), p. 554

A main goal of this thesis is to extend the pioneering work of Roedder John et al. (2006) by developing a brand concept mapping approach which also seeks to uncover the full link-by-link structure of a network of brand associations, but which does not burden respondents with the difficult cognitive task of deciding whether a link between certain brand associations actually characterizes that particular brand. Instead, as we shall show, the techniques utilized in this thesis will allow respondents to answer the much simpler and potentially much more reliable question of whether (and to what degree) a specific brand possesses each of a particular set of attributes. The technique will then utilize the dependence and independence structures that arise from these responses to derive a link-by-link description of the structure of the associated brand concept.[3, 4]

Direct versus indirect influence. The techniques described in this thesis will also help to eliminate an additional potential confound from the link elicitation procedure pioneered by Roedder John et al. (2006). Specifically, by asking respondents to decide whether a brand is characterized by a relation between two or more attributes, the structure derived via the BCM methodology may tend to conflate direct and indirect

[3] The network analysis algorithm utilized by Henderson et al. (1998), which is discussed by Roedder John et al. (2006) as an example of the analytical mapping tradition of brand concept elicitation, also utilizes consumers' evaluations of whether a brand possesses (or is characterized by) a particular attribute in order to derive the associated brand concept structure. However, as discussed earlier, this method is based on binary evaluations, and hence information about the *degree* to which a brand is characterized by a particular attribute is lost through strict dichotomization. Furthermore, the network structure is derived without reference to statistical robustness, and hence the derived structures are likely to be somewhat idiosyncratic and difficult to generalize.

[4] Structural equation modeling can also be utilized to derive a network structure from consumers' responses to ordinal or interval (typically Likert) data on the degree to which a brand possesses or is characterized by specific attributes. However, such SEM techniques require the researcher or marketer to posit an initial model and then the SEM methodology can be used to potentially validate and parameterize that possible model. The techniques described in this thesis will not require an initial model to be proposed. Furthermore, unlike most SEM based methodologies, the techniques describes in this thesis will allow one to model both linear and nonlinear dependencies among the domain variables in the model.

influences between those variables. For example, if a respondent endorses the belief that variable A directly influences variable B, and that variable B directly influences variable C, then this respondent will also quite naturally perceive that variable A has an influence on variable C. In such a case, this respondent may be tempted to directly connect variable A and variable C, even though the influence between them is actually indirect. The potential for such conflation of direct and indirect influences exists when relying on the direct elicitation of inter-association links, since such respondents may not have cognitive access to the specific link structure itself, but rather may merely sense whether one variable has an influence over another, whether through direct or indirect means.

In comparison, the techniques described in this thesis are specifically designed to discern the difference between direct and indirect influences among a given set of domain variables. In fact, such discrimination between direct and indirect influence is actually one of the core principles on which the techniques utilized here are based. As such, these methods will provide a valuable extension to the brand concept mapping technique pioneered by Roedder John et al. (2006).

Link strength interpretability. One additional area in which we would like to make a contribution is in the definition of specific link strength measures within a brand concept map. The methodology employed by the BCM technique for establishing the strength of each network link is quite direct: respondents are asked to rate the strength of each elicited link on scale of one to three (as indicated by the number of 'bars' used in that link), and then the resulting strengths of each link are averaged. However, the direct nature of this quantification procedure can also make it difficult to know what effects are incorporated into each such link strength value.

As an example, consider the excerpt shown in Figure 2, below (as extracted from a portion of the Mayo Clinic BCM network given in Figure 1). Here we see that "*Expert in treating serious illnesses*" has a direct effect upon "*World leader in new medical treatments*", with an indicated relative strength of two links. However, "*Expert in treating serious illnesses*" also has a strength-three link to "*Mayo*", which in turn has a strength-three link to "*World leader in new medical treatments*". Hence it may be unclear what the strength-two direct connection between these two associations indicates. For instance, one does not know whether this strength-two direct connection already incorporates the effects of the strength-three pathway that exists between these variables via the core Mayo brand node. Furthermore, if it does incorporate this additional pathway of influence, then it is unclear why this direct pathway only has a strength of two when these variables are connected (and hence influence each other) with a strength of three via the other pathway by which they are connected.

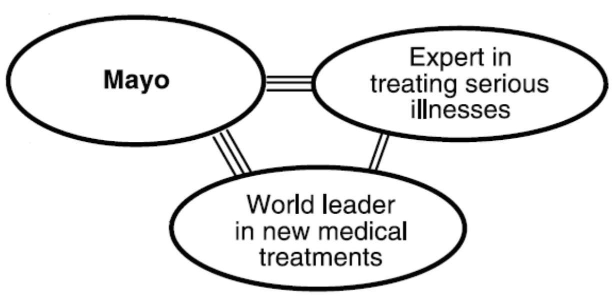


Figure 2. Excerpt from the BCM for patients' perceptions of the Mayo Clinic

Note that we are not asserting that the link strengths indicated in a BCM structure are incorrect. Since the strength assigned to each link represents an average value of the perceived strengths of that link across all the respondents whose maps contain that link, these link strength values have a *de facto* validity. What we are seeking is a delineation of what structural influences play a role in determining these link strength values. For instance, as described above, one does not know if a direct link between two attributes already reflects the contributions of all existing indirect pathways between these same two attributes. Similarly, there are known interpretability issues with directly assessed similarity judgments taken as a whole (e.g., Summers and MacKay, 1976), and the direct aggregation of individual judgments is also known to lead to possible intransitivities in the resulting aggregate link strength values (e.g., Tversky, 1969). Therefore, a potentially valuable contribution of this thesis will be the determination of a link strength measure via information-theoretic means, as well as the use of probabilistic conditioning to more precisely separate the specific contributions of each link from those of neighboring links.

1.3 Directional Relations and Marketing Constructs

As Roedder John et al. (2006) describe, the underlying cognitive model on which the BCM technique rests is that of associative networks (e.g., Anderson, 1983). On this basis, it is assumed that the links in the networks derived through the BCM technique will essentially be nondirectional entities: the variables linked together in such a structure are merely said to be associated or correlated, with no implicit directionality to their relationship. This is in keeping with the core literature cited in their work, namely

Anderson (1983a), Keller (1993), and Aaker (1996). More specifically, the BCM (along with a large number of brand construct models to date) utilizes the foundational work of Keller (1993) as guidance in establishing the theoretical core of the model. Keller, in turn, utilizes the associative network construct (Anderson and Bower, 1973; Anderson, 1983a) as the core structural assumption underlying his model of customer-based brand equity. As Keller (1993) describes these assumptions:

> … the 'associative network memory model' views semantic memory or knowledge as consisting of a set of nodes and links. . . . A node becomes a potential source of activation for the other nodes either when external information is being encoded or when internal information is retrieved from long-term memory. Activation can spread from this node to other linked nodes in memory. . . . For example, in considering a soft drink purchase, a consumer may think of Pepsi because of its strong association with the product category. Consumer knowledge most strongly linked to Pepsi should also then come to mind, such as perceptions of its taste, sugar and caffeine content, or even recalled images from a recent advertising campaign or past product experiences. . . . Consistent with an associative network memory model, brand knowledge is conceptualized as consisting of a brand node in memory to which a variety of associations are linked. (Keller, 1993, p. 2)

Interestingly, however, both Anderson's associative network model (Anderson and Bower, 1973; Anderson 1983a, 1983b) and the spreading activation model of Collins and Loftus (1975) quite often address *directional* phenomena. For example, much of the early work on learning paired associations, which played an influential role in the formulation of Anderson's model, has shown distinctly different rates of recall in the forward and reverse directions, as well as the possibility of reducing recall in one direction without affecting recall in the other direction (e.g., Keppel and Underwood, 1962; Johnston, 1967; Wolford, 1971; Anderson, 1974). Hence, in a marketing context, this would be akin to information about brand A priming the recall of brand B to a greater extent than information about brand B primes the recall of brand A (e.g., Nedungadi, 1990; Ulhaque and Bahn, 1992).

Furthermore, it is clear from the wider marketing literature that directionally asymmetric relationships exist at all levels of marketing phenomena. For example, it is well-known that consumers' understanding of a consumption situation often depends upon their perception of the directional or causal mechanism which may be generating the relevant observed or experienced attributes related to that consumption situation. For example, as Weiner (2000) points out,

> Some products lend themselves to stable attributions. For example, if I do not enjoy the taste of a breakfast cereal . . . then I will not purchase it again. After all, I expect that the next box of cereal will taste the same. (Weiner, 2000, p. 383)

However, this stable attribution relates to the perceived cause of the pleasantness or unpleasantness of the experience. For instance, as Weiner (ibid.) continues,

> Perhaps there is a chance that a hole in my tooth made the sweetness of the cereal unpleasant. This is now an unstable cause . . . so that, if I attribute my disliking to this temporary state, then I am uncertain about my future liking or disliking of

> the cereal and may try it again (i.e., I discount the properties of the cereal as the cause of my dissatisfaction). (Weiner, 2000, p. 383)

In other words, the *sense* that a consumer makes of an experience is tied to that consumer's *perceived cause* of that experience, and this relationship is certainly an asymmetric one.

As Folkes (1988) discusses, asymmetric phenomena play a central role within the process of product choice. For example, as Folkes (ibid.) describes, "many, if not most, products and services are purchased because consumers infer a causal relationship: they believe that analgesics reduce pain, deodorants improve one's social life, athletic shoes enhance performance, and so on." In other words, products and services are *efficacious*, in that their purchase or consideration involves the belief that certain of their properties will cause one or more states or outcomes to occur (or will alter the probability distribution over these states or outcomes accordingly).

One can, in fact, find directional relations at all levels of marketing phenomena. For example, as Folkes (1988) makes clear, consumers' product recommendation behavior as well as their complaint behavior are, to a large extent, based on how the consumers assign credit for good performance or assign blame for poor performance, and such assignments of credit or blame follow from the consumers' perceptions of what caused or predicted that good or poor performance. Furthermore, the ascription or attribution of causal rationale to information sources (such as spokespeople, commercials, etc.) is known to influence the perceived credibility of these sources (Wiener and Mowen, 1986, as described in Folkes, 1988).

In fact, even at the level of subcultures of consumption, directional or causal concerns often predominate. For example, as discovered by Sirsi, Ward, and Reingen (1996), the form of information which was shared among sociocognitively related subgroups was specifically *causal* information about products and product categories. Furthermore, the prominent role represented by causal or directional concerns within such subcultures makes sense, since many different subgroups can notice the co-occurrence (i.e., the undirected association) of various marketplace characteristics. However, it is the *explanation* of those co-occurrences that serves to distinguish the beliefs of one consumption subculture from another, and such explanations can often take the form of causal "if-then" rules (e.g., Hoch and Deighton, 1989).

It should come as no surprise that directional phenomena reside at the heart of so many different levels of marketing phenomena. In fact, As far back as Bartlett's classic work on human understanding and memory (Bartlett, 1932) it has been known that people's ability to comprehend a story or a script is facilitated when the elements of the script cohere with their natural understanding of the causal connection between those elements, and is inhibited when the presented flow of information does not cohere with the perceived directional relation between the relevant events. More recent studies have confirmed and extended this theme. For instance, Pennington and Hastie (1993) have shown that subjects in a mock jury experiment were far more likely to conclude that the defendant was guilty of a crime when the evidence was presented in the form of a causal story as opposed to when the exact same evidence was presented out of causal sequence. As Sloman (2005, p. 89) states when discussing these results, "Strong evidence per se does not automatically lead people to conclude guilt; the evidence must sustain an explanation. The best support for an explanation comes from a plausible causal model." In even more generality, as Sloman (ibid.) points out, legal evidence consisting of "merely statistical" facts (as opposed to causal facts) are often rejected by both judges and juries as being insufficient to prove guilt. Apparently, evidence is much more likely to be considered relevant if it is a part of a causal story or causal relation, whereas evidence that is not part of a causally relevant story is often ignored.

Furthermore, many of the heuristics and biases studied by Tversky and Kahneman (e.g., Tversky and Kahneman, 1974) can also be seen as resulting from the utilization of causal constructs in reasoning (Sloman, 2005). For example, consider Tversky's study of the "hot hand" decision heuristic (Gilovich, Vallone, and Tversky, 1985). This is the belief that individuals participating in a sport such as basketball are more likely to be successful (e.g., make their next shot) if they have already been successful on their previous few attempts. The authors' exhaustive search of relevant sports records showed that such a "hot hand" effect does not actually exist. However, the majority of players, fans, and even coaches persist in the belief that a player being "in the zone" (having a "hot hand", etc.) is indeed causal of that player making his or her next shot successfully. Apparently, the notion of a central causal mechanism which can unite the seemingly large number of variables which play a role in an athlete's performance forms a powerful enough "gestalt" to allow people to *perceive* that a causal relation exists despite objective evidence to the contrary. In fact, as Lakoff and Johnson (1980, p. 72) summarize, the causality construct plays a critical central role in people's ability to navigate their world, viz., “Our successful functioning in the world involves the application of the concept of causation to ever new domains of activity – through intention, planning, drawing inferences, etc.” In fact, as these authors also point out, the experiential gestalt of causality is one of the “ultimate building blocks of meaning” (Lakoff and Johnson, ibid., p. 69).

1.4 Mechanisms Underlying Asymmetric Relations

Given the central role played by directional relationships in people’s everyday experience, it stands to reason that any model of natural reasoning mechanisms would have to be able to explain such directional phenomena. Within the associative network model of human reasoning, the key structural element which is hypothesized to generate directional asymmetry is known as the *fan effect,* which corresponds to the assumption that multiple associational links lead out of each node much like the shape of a fan (Anderson, 1974, 1983a). Because nodal activation spreads along these links, the activation of a particular node will be diminished more quickly if a greater number of associations lead out of that node (i.e., the rate of extinction of a node’s activation is proportional to the size of that node’s fan). Stated another way, given a fixed amount of nodal activation, the degree or amount of that activation which traverses any one link leading out of the activated node will be inversely related to the number of other links in the network which share that same source node. Therefore, if nodes A and B within an associative network are connected, and node A has a larger ‘fan’ structure than node B (i.e., more associative links lead out of node A than node B), then activation of node A will have less of an effect on node B than an equally strong activation at node B has upon node A (where the ‘effect’ on a node is understood in the associative network sense, i.e., as the probability that the destination node will become activated given that the source node has been activated).

The fan effect has been experimentally verified in multiple studies (e.g., Anderson, 1974; Lewis and Anderson, 1976; Reder and Ross, 1983; Reder and Wible, 1984; Anderson and Reder, 1987; Schreiber and Nelson, 1998). In fact, much of Anderson’s seminal treatise on associative networks as the foundation of cognition (Anderson, 1983a) consists of verifying and quantifying this fan effect within multiple different cognitive tasks and domains.[5]

Thus there are two distinct but related levels of analysis at which associative network theory operates. At the most basic functional level, there is simply a collection of undifferentiated, potentially bidirectional structural links connecting various nodes, much as a set of neuronal connections are assumed to link various neural clusters in the brain (Anderson, 1983a, p. 86-87). However, once the *overall topology* of the multiple links in an associative network is taken into account, the fan effect may then become

[5] Many examples of the fan effect examined by Anderson (1983a) consisted of concept pairs which span a subordinate-superordinate category dimension (i.e., one member of each pair was often clearly superordinate to the other). Hence, as classically described by Anderson (1983a), the fan effect may be said to be an explanation of the possible neural structures which may encode the perception or understanding of subordinate-superordinate relationships, and as shown by Anderson (ibid.), the associative network representation of such relationships clearly exhibit asymmetric priming and facilitation effects. Further, Anderson demonstrated that such directional differences in priming and recall follow from differences in the number of possible associative links emanating from the concepts at each level of such superordinate-subordinate hierarchies (i.e., from the size of the ‘fan effect’ differential between the associatively linked concepts).

operative and hence the overall association between variables that results from the network will frequently exhibit a clear directionality.[6]

Interestingly, more recent investigations into the boundary conditions of the fan effect have shown that when the additional information presented to subjects in an association task is schematically or causally connected, the ‘fan effect’ is frequently eliminated (e.g., Radvansky, 1999; Gomez-Ariza and Bajo, 2003). In essence, this research shows that the existence of a causal connection between the additional presented facts (or a thematic coherence among those facts that would allow a causal representation) provides the conditions under which respondents are able to mentally ‘group’ or ‘coalesce’ the multitude of simple bidirectional links emanating from a source node into a single directional (causal) link, thereby removing the ‘fanning out effect’ of the multiple links emanating from that source node. Hence, in order to ‘replace’ a set of undifferentiated links (whose complex topology can result in a directional dependence between the linked variables) with a simpler connection structure, that replacement structure typically must itself be causal in nature.

The notion of the associative ‘fan effect’ has also been demonstrated within marketing phenomena. For instance, Lei, Dawar, and Lemmink (2008) created fictitious brands so that the number of associations linked to each brand could be experimentally controlled, and found that the directional strength of association between these brands

[6] Additionally, some models of associative network structures place either *a priori* or variable ‘strengths’ or ‘criterialities’ on the various links within the network (e.g., Collins and Quillian, 1972), and it may be possible to create an asymmetry through specifying different criterialities in each direction. This argument essentially parallels the ‘fan size’ argument described above. (In fact, such differing criterialities may be derivable from differing fan effect sizes for each variable in the connection.)

followed the precise pattern predicted by the associative ‘fan effect’, namely that the priming of a brand with a greater number of associations had less of an effect on the recall of a related brand with a smaller number of associations than did the converse condition of priming a brand with a smaller number of associations. Further, by manipulating the number of shared versus distinctive attributes among the brands in this portfolio, Lei et al. (ibid.) were able to rule out Tversky’s well-known contrast model (Tversky, 1977; Tversky and Gati, 1978) as a possible explanation of these directional effects. Furthermore, these investigators were able to show that the degree to which negative information (i.e., a ‘brand crisis’) at one brand within a portfolio was able to affect evaluations of other brands in that portfolio precisely followed from the relative directional strengths of the links between these brands.

One can also posit the existence of an associative fan effect within many of the directional phenomena known in marketing. For example, consider the illustrative example provided by Holden and Lutz (1992) of an asymmetric dependence between the Budweiser brand and the Superbowl usage situation. As they describe, when primed with the usage situation (i.e., when thinking about something to drink when watching the Superbowl), the Budweiser brand may come to mind, but thinking about the brand is much less likely to activate that usage situation. Hence there is a clear asymmetric dependence between the brand (Budweiser) and the usage situation (something to drink during the Superbowl). In terms of the underlying fan effect within this situation, it is clear that the Budweiser brand has one of the richest and most varied brand images of any beer brand in the United States, and it is quite likely that a typical American male respondent would be able to name vastly more associations to the probe ‘Budweiser’ than to the probe ‘something to drink during the Superbowl’. Thus ‘Budweiser’ is likely to have a larger (and more complex) fan structure, and hence by the fan theory of associative networks, activation of the node or concept ‘something to drink during the Superbowl’ would have a stronger effect on the level of activation of ‘Budweiser’ than the activation of ‘Budweiser’ would have on ‘something to drink during the Superbowl’.

As the above example illustrates, the relation between many of the concepts within consumers’ brand concepts may in fact have a directional nature. Furthermore, such directionality has been utilized in the marketing management and consumer behavior literatures. For example, Farquhar and Herr (1993) stress that *brand building* activities, which focus on strengthening the directional relationship from the focal brand to its many brand associations are different from *brand leveraging* activities, which focus on strengthening the directional relationship stemming from a brand’s associations and terminating at the brand itself.[7] As further evidence of this directional asymmetry, Farquhar and Herr (ibid.) also distinguish between a category dominant brand (which has a very high tendency to be named or recalled once the superordinate brand category is primed), versus an instance dominant brand (for which priming of the brand has a high likelihood of resulting in the brand category being named or recalled). For instance, Farquhar and Herr (1993) describe how the brand ‘Nike’ is both category dominant (if consumers are asked to name a sneaker brand, there is a high probability they will name

[7] As Farquhar and Herr (1993) describe, the managerial concepts of brand building and brand leveraging are essentially derived from the concepts of *instance dominance* and *category dominance* respectively (e.g., Barsalou 1983, 1985).

‘Nike’) and at the same time is also instance dominant (if consumers are primed with the ‘Nike’ brand, there is a high probability that this will activate the sneaker category). However, as these authors show, these two forms of directional relation do not necessarily go hand-in-hand. For instance, these authors show that ‘Scope’ is more category dominant than ‘Listerine’, but ‘Listerine’ is more instance dominant than ‘Scope’. Hence, as demonstrated by Farquhar and Herr (1993), brand-category relationships in marketing are typically asymmetric phenomena.

Clearly, based on the evidence discussed, the links within consumers’ brand concepts, as well as between consumers’ perceptions of multiple brands within a portfolio, can be asymmetric. Furthermore, based on evidence from the psychology and marketing literatures, the degree of asymmetry of these link strengths does seem to follow from the predictions of the ‘fan effect’ as initially proposed within the associative networks literature. Hence, consumers’ brand maps may in fact be better modeled by asymmetric or directed networks as opposed to presuming that all associative links must be nondirectional by nature.[8] Interestingly, the pioneers of the BCM methodology are distinctly aware of these considerations, and these authors specifically mention the idea of link directionality as a potential worthwhile extension of the BCM methodology as it currently exists. Specifically, Roedder John et al. (2006) state:

> … it would be useful to incorporate procedures into the BCM to assess the nature of relationships between associations, that is, whether it is causal, correlational, or

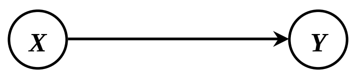
[8] There is actually nothing in Keller (1993) which specifically states that the links in an associative network model of a brand concept *must* be nondirectional. Rather, for the purposes of developing his consumer-based model of brand equity, it is simply quite likely that Keller (ibid.) saw no overt need to posit any form of directionality to the links in the derived network.

something else. Although we can speculate about the relationships shown in the consensus brand maps, we have not yet developed a technique for doing so on an objective basis. For example, it seems clear that perceptions of Mayo Clinic as 'treats famous people around the world' cause people to believe that Mayo Clinic is 'known worldwide.' However, being a 'leader in cancer research' could be an instance of being a 'leader in medical research,' or one of these associations could be driving (causing) the other. We believe that procedures similar to those used in understanding causal reasoning chains (see Sirsi, Ward, and Reingen 1996) could be incorporated into the mapping stage of the BCM to provide information about brand association relationships. (Roedder John, et al., 2006, p. 563)

1.5 Causality, Diagnosticity, and Intervention

In order to further pursue the suggestion of Roedder John et al. (2006) that it may be quite plausible for the links in a brand concept map to be causal or directional in nature, we must examine what such a link entails for a network of variables or brand associations, and how it may differ from a purely associative link between those variables. At first glance, it may seem that if there is a causal or directional link from variable X to variable Y (denoted as $X \rightarrow Y$) within a brand concept network, then knowing something about the state of variable X would imply something about the state of variable Y, but not vice-versa. However, this initial supposition is not actually true.

To clarify this point a bit further, let us assume that we have such a directional relationship between two binary variables X and Y. In such a case, the distributions related to this bivariate relationship would be of the form given in Figure 3, below. As indicated by this figure, if we observe that variable X happens to have value x_1, the distribution of variable Y would essentially be 'collapsed' to the first row of the conditional distribution $p_{Y|X}$ as given in the right-hand data table within Figure 3. In other words, this knowledge about the value of variable X would alter the distribution of variable Y from whatever its initial values were, so that variable Y would now have the values indicated in the x_1 row of the $p_{Y|X}$ data table.



p_X	x_1	x_2
	λ	$1-\lambda$

$p_{Y|X}$	y_1	y_2
x_1	α	$1-\alpha$
x_2	β	$1-\beta$

Figure 3. Directed relation $X \rightarrow Y$ involving two binary variables

Now consider what would occur within the directional structure of Figure 3 if we learn that variable Y happens to have the value y_1. Contrary to our initial assumption that the 'flow of causality' from X to Y would preclude knowledge of variable Y from having an effect on variable X, we would actually see a definite effect on variable X, and this effect can be quantified by Bayes' theorem. Specifically, the value of $p_X(x_1)$ will be updated from λ to $\frac{p_{X|Y}(y_1|x_1) \cdot p_X(x_1)}{p_Y(y_1)}$, which in this instance can be computed as $\frac{\alpha\,\lambda}{\alpha\,\lambda\; +\; \beta\,(1-\lambda)}$. Hence, unless either $\lambda = 0$ or $\lambda = 1$ (which are 'degenerate' cases for the distribution of the variable X) or $\alpha = \beta$ (which would mean that variable Y is "accidentally" perfectly independent of variable X to begin with, in which case the model $X \rightarrow Y$ that we are utilizing would no longer be valid), we will have that an observed value of variable Y will provide diagnostic information about the value of variable X, and we will therefore see bidirectional influence between these two variables.

If, as we have shown, Bayes' theorem allows for diagnostic influence from the 'effect' or 'consequent' variable to the predictive or causal variable within that pair, one may wonder what the difference is between this directional (or asymmetric) relationship and a correlational (or symmetric) relation. As it turns out, the difference comes about when we *act* to *set* the value of one of the variables rather than merely *observing* its value.

More specifically, within a symmetric (correlational) relationship, the means by which the value of either X or Y is established (i.e., whether the variable is set to a specific value or is observed to have that value) has no influence upon the determination of the effect that this variable has upon the other variable within the correlated pair. However, when the relationship between X and Y is directional, such as is the case in the scenario exhibited in Figure 3, each direction of the relationship responds differently to an intervention to set the value of one of these variables (Lauritzen, 1999). For instance, as long as variable Y is not, by chance, independent of variable X (i.e., assuming that the rare coincidence of $\alpha = \beta$ does not occur)[9], then acting to fix the value of variable X

[9] This assumption, that the two variables are not independent 'by coincidence only' (i.e., that the causal diagram $X \rightarrow Y$ is valid) is called the *causal faithfulness* assumption, and will be discussed at greater length later in the thesis.

will still have the same effect as observing the value of X, namely, such knowledge will serve to 'collapse' the distribution of variable Y to one particular row of its conditional distribution.[10]

In contrast, acting to set the value of variable Y will have no effect on the distribution of variable X in a directional system such as the one in Figure 3 *because the diagnostic influence that runs from variable Y to variable X is only operative when the value of variable Y actually indicates (or 'diagnoses') something about the state of variable X*. When the value of Y is established through an intervention rather than through the normal causal mechanism running from variable X to variable Y, then knowing the value of Y tells us nothing about the possible states of variable X.

As an example of this phenomenon[11], consider the fact that observing a low reading on a barometer is certainly diagnostic of the fact that rain is likely. Thus even though a low reading on the barometer does not *cause* the atmospheric conditions to favor rain, we can still utilize the low reading as an indicator of those atmospheric conditions. This is a *diagnostic* relation (and not a causal one) because the conclusion that the low reading indicates rain-like conditions runs in the opposite direction to the underlying directional or causal mechanism, namely: *rain-like conditions → low barometer reading*). However, despite the fact that this diagnostic reasoning pattern runs in the opposite direction to the underlying causal mechanism, we still strongly believe in its validity because we believe in the validity of the causal mechanism itself (i.e., *rain-*

[10] In essence, the variable Y does not "care" who or what caused variable X to have the value that it does. Variable Y simply "senses" the value that its cause (variable X) currently has, and responds accordingly.

[11] This is an extension of an example described in Pearl (2000, p. 111).

like conditions → *low barometer reading*), and we are quite used to 'reversing' trusted causal mechanisms (via a an intuitive version of Bayes' theorem).

On the other hand, if we were to open up the barometer and *set* the reading to be low, we would certainly not believe that this reading was indicative of a high probability of rain. This negative conclusion of course stems from the fact that because we intervened to *set* the value of the effect variable (i.e., the barometer reading), we clearly no longer regard the value of the effect variable as being indicative of the distribution of the causal variable, and hence we are likely to dismiss any possible effect in the diagnostic direction. Interestingly, if we could somehow 'seed' the clouds in order to establish particular barometric conditions, we would quite readily believe that this intervention would indeed lower the barometer reading, even though we intervened to 'set' those atmospheric conditions.

Thus, as this example shows, we typically trust directional mechanisms in the causal direction no matter whether the level of the causal variable has been observed or has been manipulated to be at a certain level or value through an intervention. On the other hand, we only trust causal mechanisms in the diagnostic direction if we *observe* the value of the effect variable, but not if we set that value through some intervention. The fact that these reasoning patterns appear so intuitive (i.e., of such a "second nature" to us) reveals how prevalent and central these directional reasoning patterns really are within our conception of the world.

In summary, we have shown that a causal link between two variables will be apparent in an interventional scenario, but not necessarily in an observational one. The value of the causal variable will always influence the distribution of the effect variable, whether we intervene to set the causal variable's value, or whether we passively observe its value. However, the value of the effect variable will influence the distribution of its cause only when we observe the value of the effect (and diagnostically reason from the value of the effect to the distribution of the cause), but not when we intervene to set its value. Therefore the asymmetry that Roedder John et al. (2006) discuss as potentially providing a deeper explanation of link structure within a brand concept map will typically arise under conditions in which we *set* the values of certain variables, rather than when we merely *observe* their values, and will be manifested in consumers' sensitivity to differences between the predictive and diagnostic directions of the resulting links.

Note that since we are analyzing consumer belief structures, the idea of fixing or setting the values of certain variables with in a consumption domain can also correspond to consumers' *beliefs* about what *would happen* if they *were* to set these values. In other words, the 'fixing' of values in this case can just as easily derive from counterfactual thoughts about what is likely to happen if one were to take a particular action as it is to derive from actual actions undertaken by consumers or firms in the marketplace. In either the enacted or counterfactual scenario, we can expect that many of the links between variables in the belief network may be directional in nature. Furthermore, as discussed above, the asymmetric response of such a directed network to interventions allows one to regard a directional network as an oracle for determining the likely effects of either real or counterfactually imagined choices among the brand characteristics that exist within the consumer's brand concept.

1.6 Directed Structures, Part I: Common Cause and Common Effect Structures

As we have seen in the previous section, directional networks are sensitive to the difference between observing a fact versus intervening to establish that fact, whereas nondirectional networks do not necessarily possess such sensitivity.[12] In this section we will see that the additional expressivity available when one allows the links within a brand concept network to be directed also allows one to distinguish several different variations of the relationship between triples of network variables, whereas in an undirected network all of these variations would collapse into a single nondirectional structure. Furthermore, as it turns out, *the difference between these directed triples of variables is precisely what provides directed networks with their rich implicational structure and semantic meaning.*

To be more specific, consider a chain structure consisting of three variables, say, A, B, and C. In an undirected network, such a chain can have only one basic configuration[13], namely A – B – C. However, within a directed network, this chain-like structure can actually have four different variations, namely the *causal chain* structure $A \rightarrow B \rightarrow C$, the *diagnostic chain* structure $A \leftarrow B \leftarrow C$, the *common cause* structure $A \leftarrow B \rightarrow C$, and the *common-effect* structure $A \rightarrow B \leftarrow C$. As it turns out, the

[12] This can also give rise to different reasoning patterns in the predictive versus the diagnostic direction (e.g., Tversky and Kahneman, 1980; Pearl, 1988), which is another distinction that is expressible (and modelable) within directed networks, but which is difficult to represent within nondirectional networks.

[13] In this discussion, we are ignoring possible differences in the 'naming' of the three variables involved. For example, we are considering the undirected chain A – B – C to represent the same basic network structure as B – A – C , etc., since these have the same network topology and differ only in the names given to the three variables occupying the three positions within that structure.

semantic differences between these four directional variants is actually quite profound (e.g., Lauritzen, 1999; Pearl, 2000).

As an example of the depth of the implicational differences between each of these directed three-variable substructures, consider a typical situation in which a consumer utilizes a product review in order to ascertain whether or not he or she is likely to be satisfied with the purchase of a particular product under consideration. Of course, we could represent this situation with an undirected network as shown in Figure 4a, but the directed representation given in Figure 4b would actually be more accurate since it reveals that the product quality directly influences both the review and the consumer's satisfaction with the product, but that neither the review nor the consumer's satisfaction directly influence the product quality.[14]

A consumer who holds the directed cognitive model shown in Figure 4b will be able to use the status of the product review to diagnostically infer the likely quality of the product in question, and can then infer whether he or she will be satisfied with that product. In other words, within this directed structure, information flows from the status of the product review to the determination of the consumer's likely degree of satisfaction with the product.

[14] In both Figure 4a and Figure 4b, we are ignoring the possibility that the customer will derive product satisfaction from the review itself. Rather, we are assuming that the consumer only cares about the underlying quality of the product, and is simply using the review as an informational tool to learn about that underlying product quality. (In this simple example, we are also ignoring any potential 'second-order' effects by which a positive or negative review or could 'feed back' to affect product quality through managerial response to that review, or by which managerial knowledge of overall customer satisfaction alters current product quality.)

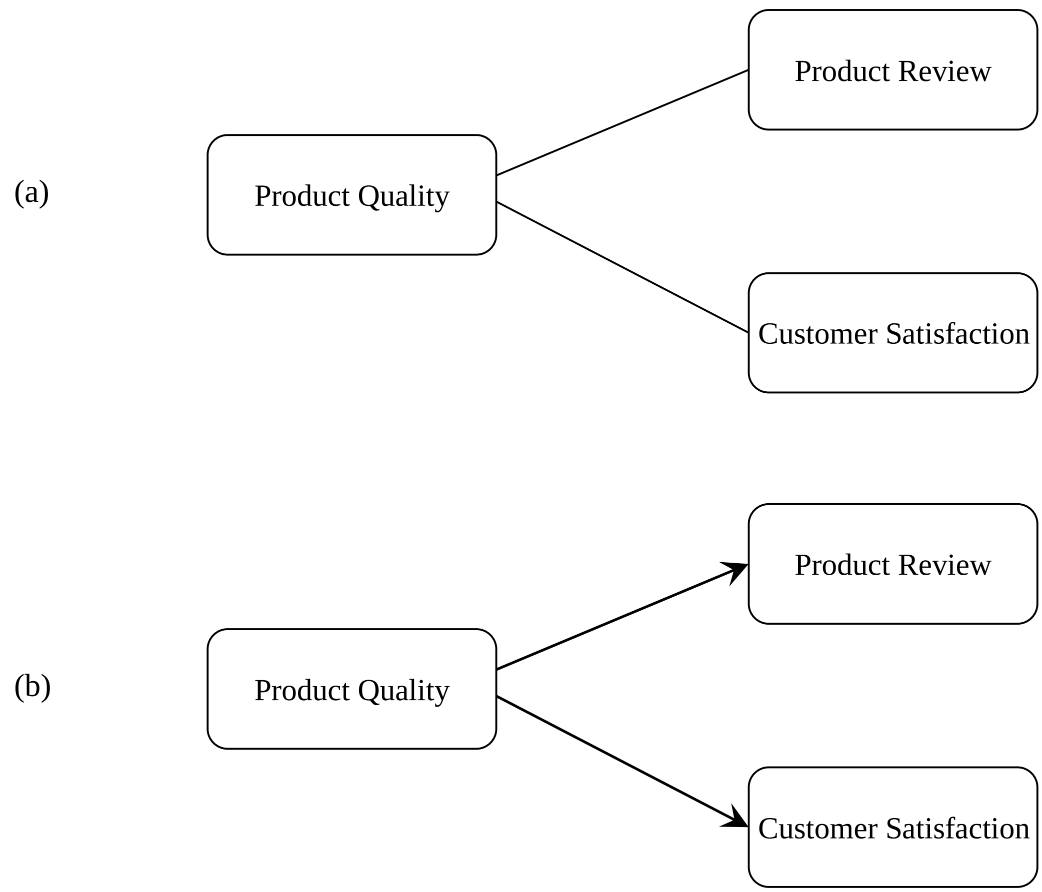


Figure 4. Alternative models for a product review scenario

However, suppose that this consumer sets the value of the ‘Product Quality’ variable by only incorporating items of a known quality level into his or her consideration set. In such a situation, the product review would become irrelevant to the purchase decision, since the consumer can ascertain the quality of the product through this other mechanism.[15] Hence, once the value of the central variable in this common cause structure is fixed (in this case through an intervention, either ‘real’ or counterfactually simulated) by the consumer, the flow of information from the ‘Product Review’ variable to the ‘Customer Satisfaction’ variable will be blocked. In probabilistic terms, knowledge of the ‘Product Quality’ variable “blocks” the passage of information through this structure, rendering the ‘Product Review’ and ‘Customer Satisfaction’ conditionally independent of each other: *Customer Satisfaction* $\perp\!\!\!\perp$ *Product Review* | *Product Quality*. Similarly, note that one can also block the passage of information through this structure via a direct observation of the product’s quality, say, through direct inspection of the product by the consumer, or through a product trial, etc. In other words, the information flow through this structure can be blocked either by an *intervention* to set the level of the ‘Product Quality’ variable, or via an *observation* of the product’s quality, and in either case, we would have that the two terminal variables in the structure (namely ‘Customer Satisfaction’ and ‘Product Review’) become independent when knowledge of the central variable (‘Product Quality’) becomes available.

Now consider a slight variation on this three-variable scenario. For instance, suppose that a consumer is shopping for a new stereo or television, etc., and that he or she holds the directional belief structure shown in Figure 5. Note that if this were an *undirected* structure, it would have the *same* overall topology as each of the structures shown in Figure 4, namely a central variable linked to each of two terminal variables, with no direct connection between those terminal variables. However, what is quite surprising is that despite its topological similarity to the structures in Figure 4, the

[15] In this basic model, we are assuming that the product review would not add any additional degree of confidence in the product above and beyond the level of confidence that the consumer has already established via the incorporation of products with a specific known level of quality into their consideration set. (Such a more complex scenario would require an expansion of this basic model.)

directional variation shown in Figure 5 will actually have informational properties that are completely different than what we saw in the previous example.

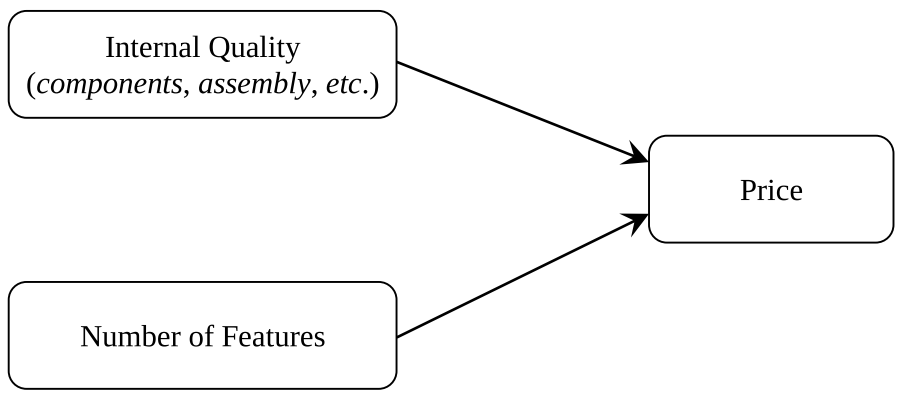


Figure 5. Common effect structure

To derive the informational semantics of this form of directed structure, consider what a customer might think if he or she encountered a potential choice of stereo which had a great many features. If this consumer did not know the price level of the stereo, he or she would have no way of knowing whether the high number of features was indicative of a high quality or a low quality product.[16] However, if the consumer *does*

[16] Some consumers may perceive that certain brands are better than others at increasing feature counts without sacrificing quality. However, to keep this model as simple as possible, we are assuming that the consumers in the model do not have an opinion one way or the other on this particular issue. Furthermore, for expository purposes, we are disregarding other signals of internal quality (such as channel exclusivity, retailer reputation, etc.).

know the price level of the product, then the number of features offered by that product certainly provides information about the product's quality level. Specifically, if the price is quite low, then the product's high number of features would almost certainly indicate to the consumer that this particular stereo was likely to have a low level of internal quality. Thus without knowledge of the value of the central variable in this directional structure, one of the terminal variables provides no information about the other, and hence these two variables are said to be unconditionally independent, i.e., $\textit{Internal Quality} \perp\!\!\!\perp \textit{Number of Features} \mid \varnothing$. Furthermore, once the price level becomes known (either through observing that price level, or through setting that price level via limiting one's choice of products to incorporate into a consideration set), the number of features and the perceived quality of the stereo become linked, and hence we have that these variables are *conditionally dependent* on each other given the price level, i.e., $\textit{Internal Quality} \not\perp\!\!\!\perp \textit{Number of Features} \mid \textit{Price}$.

To summarize these findings, let us symbolically denote the variables in each structure examined by A, B, and C. Thus, in the product review example of Figure 4b (which we can symbolically denote by $A \leftarrow B \rightarrow C$), we saw that this structure possessed the property that the two terminal variables (i.e., the product quality and the consumer's likely degree of satisfaction with the product) initially provided information about one another, but became independent of one another once the central variable in the structure (namely the product's quality level, or symbolically variable "B" in this structure) was known or was fixed through either real or counterfactual intervention. On the other hand, the stereo purchase example of Figure 5 (which we can symbolically denote by

$A \rightarrow B \leftarrow C$) had the *opposite* set of independence and dependence properties, namely the two terminal variables (number of features and internal quality) initially do not provide information about one another, but do become related once the central variable (i.e., the price, or variable "B" in the schematic representation) is known or fixed through either real or counterfactual intervention.[17]

We can state the above findings more succinctly by saying that the *common cause structure* $A \leftarrow B \rightarrow C$ possesses marginal *dependence* of A and C, but conditional *independence* of A and C once variable B becomes known (or is fixed through intervention). On the other hand, the *common effect structure* $A \rightarrow B \leftarrow C$ possesses marginal *independence* of variables A and C, but conditional *dependence* of these variables once B is known (or set through an intervention). Stated even more simply, within a *common cause* structure $(A \leftarrow B \rightarrow C)$, the central variable *blocks* any communication or information flow between the two terminal variables, while in the *common effect* structure $(A \rightarrow B \leftarrow C)$ knowledge of the value of the central variable is *required* in order to permit information flow between the two terminal variables.[18] Of course, note that had these two phenomena been modeled as undirected structures, both models would have been identical (namely $A – B – C$), and all of the rich expressive power that was gained through these different directional representations would have been lost.

[17] Once again, by "real or counterfactual intervention", we mean that the consumer either limits his or her consideration set to products containing a specific level or value of the variable in question, or counterfactually considers what would happen if he or she considered only products with a specific value of that variable.

[18] Proofs of these differences in the informational dynamics of the common cause and common effect structures are provided in the Appendix.

Lastly, it should be pointed out that within a common effect (or 'collider') structure ($A \rightarrow B \leftarrow C$), we have seen that confirmation of one of the possible causes of the common effect has no influence on the other possible cause until the level of the common effect is known, at which time the confirmation of one possible cause reduces the belief in the other possible cause of that common effect. For example, as described earlier, both the number of features of a stereo and its internal quality have a direct influence on the price of that stereo. However, once the price of the stereo is observed (or is fixed through intervention), any increase in the perceived number of features of that stereo will tend to decrease consumers' perceptions of the possible internal quality of that stereo. This reasoning pattern is often called *explaining away* (or '*intercausal reasoning*'), and may be used to provide quantitative predictions of the degree to which changes in the probability of one or more causes of a common effect reduces the perceived likelihood of the other causes of that effect (Wellman and Henrion, 1993; Pearl, 2000). Further, note that this hallmark property of the common effect structure is also known by the names *causal attribution* and *causal discounting* within the social psychology literature (e.g., Kelley, 1973), since it models the process by which the attribution of causal strength to one possible cause of a common effect has the result of diminishing the causal strength associated with other possible causes of that same effect.

1.7 Directed Structures, Part II: Causal and Diagnostic Reasoning Chains

There are, of course, two additional directional structures (besides $A \leftarrow B \rightarrow C$ and $A \rightarrow B \leftarrow C$) which can exist within a chain of three variables: namely the *causal reasoning chain* $A \rightarrow B \rightarrow C$ and the *diagnostic reasoning chain* $A \leftarrow B \leftarrow C$. Interestingly, as proven in the Appendix, these two directional structures actually share the same informational dynamics as the common cause structure ($A \leftarrow B \rightarrow C$), namely the two terminal variables are dependent upon one another, but become independent once the value of the central variable (B) is either observed or is set through intervention. However, even without going through the specifics of the proof, we can observe this behavior by analyzing some very common marketing phenomena.

For example, consider the causal reasoning chain shown in Figure 6, which models a possible consumer belief system for pain relievers. For a consumer holding this belief structure, the brand of a pain reliever and the belief that this particular pain reliever will alleviate their headache are probabilistically dependent upon each other, but once it is known whether or not the pain reliever contains a particular active ingredient that the consumer believes is effective at headache reduction, the brand name of that medication becomes irrelevant to that consumer's belief in whether or not the medication will relieve their headache.[19] Hence in such a structure, we have that the two terminal variables in the chain are rendered independent of one another once the value of the central variable

[19] In this model, we are assuming that the consumer does not believe in any additional pathways from the brand to headache relief that do not pass through the '*active ingredients*' node. For example, we are assuming that such a consumer would not believe that some brands' non-active ingredients (such as their 'fillers' or pill coatings, etc.) or even their location of manufacture would influence the medication's headache-reducing capacity. (If the consumer were to hold such additional beliefs, we would need to add additional pathways to the model accordingly.)

in the chain becomes known (either through observation of the ingredients in a potential pain reliever choice, or through intervening to set the level of the central variable via limiting one's consideration set to just those products which contain a certain level of a particular active ingredient). In other words, within this causal reasoning chain structure, we have that $(\textit{headache relief} \perp\!\!\!\perp \textit{brand}) \mid (\textit{active ingredients})$. Of course, as this example illustrates, we have that in any causal reasoning chain $A \to B \to C$, the two terminal variables A and C will be dependent in probability ($A \not\perp\!\!\!\perp C$), but are rendered independent once the value of B is observed or established through intervention, i.e., the conditional independence relation $(A \perp\!\!\!\perp C) \mid B$ holds for this structure.

Figure 6. Causal reasoning chain model for pain relievers

Now consider a scenario in which a consumer observes the 'outcome' of a causal chain structure, and wishes to utilize this information to reason about the state of the initial variable within that structure. Such a case would occur, for instance, if the consumer in the previous example observed that a particular medication was able to relieve his or her headache, and was using this information to reason about the probability that the medication is likely to come from a particular brand.[20] This example is illustrated in Figure 7 (in which the direction of inference is shown via dashed arrows). Since this reasoning problem runs in the opposite direction to the predictive or causal direction that exists in the structure, this is termed a *diagnostic reasoning chain*.

Figure 7. Diagnostic reasoning chain model for pain relievers

To discuss the diagnostic reasoning pattern in this example, we have to assume that the consumer in question has two characteristics. Firstly, we must assume that this consumer holds the corresponding *forward* (or 'predictive') reasoning pattern (i.e., that the predictive relations from '*brand*' to '*active ingredients'* and from '*active ingredients'* to '*headache relief*' are valid). Secondly, in order for this consumer to need to engage in diagnostic reasoning, we would have to assume that he or she is not certain of the medication's brand, since otherwise he or she would have no need to reason diagnostically in the first place. Since we are assuming that the consumer in this example is not certain of the medication's brand, then he or she will simply have a belief

[20] For instance, we can suppose that this consumer took a pill from a collection of mixed medications that he or she had with them, or borrowed a pain reliever from a co-worker, etc.

distribution over the possible brands that this medication could be from. In other words, this consumer would essentially have a prior probability that the medication comes from each particular possible brand.

Diagnostic reasoning would then occur in this situation once the consumer learns of the headache-reducing capacity of this particular medication (for instance, by either taking the medication or by receiving reports from other sources about its headache-reducing power, etc.). Once this information about the medication's headache-reducing power becomes known, the consumer is likely to *revise* his or her beliefs about the probability that the medication comes from each of the different possible brands.[21] *This belief revision represents the diagnostic reasoning process*. In essence, information in this diagnostic reasoning chain 'flows' from the variable '*headache relief*' to the variable '*brand*', thereby modifying the consumer's prior probability distribution over the various possible brands that the medication could have come from.

As is proven in the Appendix, the flow of (diagnostic) information from the 'outcome' (or 'effect') variable *headache relief* towards the 'predictive' (or 'causal') variable (*Brand*) can proceed freely as long as the intermediate variable (*active ingredients*) is not observed or established at any particular value. This conclusion is logical for a consumer who holds a belief pattern such as this, since once such a consumer knows whether or not the pain reliever contains a specific active ingredient,

[21] For example, if the medication is found to be quite effective at reducing the consumer's headache symptoms, then this consumer would likely increase the probability that he or she places on the medication having come from an ibuprofen-containing brand (and concomitantly reduce the probability that he or she places on the medication coming from a non-ibuprofen-containing brand).

that consumer can make a conclusion about the likely brand of the medication without needing to know about its headache-reducing capacity.[22]

In essence, a diagnostic reasoning chain appears quite similar to a causal reasoning chain, in that both have chain configurations, both have the same conditional independence property (namely, $A \perp\!\!\!\perp C \mid B$), and in both structures the flow of information proceeds from one variable to a second variable, and then from that second variable to a third variable (i.e., both directed links point in the same direction). Therefore, the reason for distinguishing these two structures from one another lies not within their overall topology or conditional independence properties (which is the same in both cases), but rather because of the differences in the means by which information is updated in each structure.

Within a causal chain such as the example in Figure 6, information is propagated in the same direction as the predictive or causal mechanism, and hence updates to either the terminal variable or the intermediate variable in this structure can be normatively computed simply by applying the distribution of that variable in the chain conditional on the observed (or fixed) value of the prior variable in the chain. For instance, suppose that

[22] One might hypothesize that if the medication failed to alleviate the consumer's headache, that this would provide additional information about the possible brand. However, according to the belief system represented in this model, such an observation would essentially already be incorporated into the consumer's conditional probability of headache relief given various levels of active ingredients (i.e., it would already have been incorporated into the consumer's beliefs about the effectiveness of various active ingredients). Therefore, such an observation (that the medication did or did not reduce the consumer's headache) could alter the consumer's beliefs about the possible active ingredient combinations that the medication may have. However, once we assume that we ***know*** the particular active ingredient configuration of the medication, the additional headache reduction observation is rendered irrelevant to the deduction about the possible brand, since all it tells us is something about the possible active ingredient levels, which is something we would already know for certain (since we observed those active ingredient levels or set them through the construction of our consideration set).

a consumer perceives there to be four levels of headache relief (say: excellent, average, minimal, and none), and five values of the ‘active ingredients’ variable (say: low ibuprofen, high ibuprofen, low acetaminophen, high acetaminophen, and ‘other’). Furthermore, suppose that the consumer’s initial distribution of the headache relief variable is $(\alpha, \beta, \gamma, 1 - \alpha - \beta - \gamma)$. Now suppose that this consumer either observes or fixes the value of the ‘active ingredients’ variable to be ‘low ibuprofen’ (L.I. for short). This consumer’s updated distribution (conditional on *active ingredients* = L.I.) would become: (P(relief=exc.| LI), P(relief=avg. | LI), P(relief=min,| LI), P(relief=none | LI)), and these values would occupy the row in the conditional distribution table for the *headache relief* variable which is indicated by the value *active ingredients* = L.I.

On the other hand, if a consumer observed the value of the *headache relief* variable and wanted to reason ‘back’ (i.e., diagnostically) to the likely value of the active ingredients variable, this diagnostic reasoning process would normatively proceed via Bayes’ theorem (rather than as a simple predictive conditional probability, as in the previous example). For example, if the consumer’s level of headache relief was average (denoted by *rel* = *avg*), then his or her distribution of the *active ingredients* variable would (normatively) become updated from its prior distribution to a new distribution with parameter values such as:

$$P(ib = high \mid rel = avg) \ = \ \frac{P(rel = avg \mid ib = high) \ * \ P(ib = high)}{\sum_j P(rel = avg \mid act\ ingred_{\,j})} \qquad (1)$$

where “$ib = high$” denotes a high level of ibuprofen, and $act\ ingred_j$ denotes the j^{th} level of the *active ingredients* variable (and where $ib = high$ corresponds to one possible value of $act\ ingred_j$). Of course, equation (1) only shows the updated value for the probability placed on *ibuprofen* = *high* by this consumer. There would also be equivalent calculations for updating his or her perceived probabilities for the other possible values of the *active ingredients* variable once the consumer has experienced a particular level of headache relief.

In summary, we have seen that despite the differing process by which information is propagated in causal reasoning chains versus diagnostic reasoning chains, the overall conditional independence properties of these two structures are nonetheless the same. Specifically, given a causal reasoning chain $A \rightarrow B \rightarrow C$ and a diagnostic reasoning chain $A \leftarrow B \leftarrow C$ (for which $A \rightarrow B \rightarrow C$ is the underlying ‘causal’ direction), we will have that in both cases, variables A and C will be rendered independent once the value of variable B is known or is fixed through an intervention, i.e., $(A \perp\!\!\!\perp C) \mid B$ in both cases.

1.8 Markov Equivalence

As demonstrated in the previous section (and as proven in the Appendix), both the causal reasoning chain structure $A \rightarrow B \rightarrow C$ and a diagnostic reasoning chain structure $A \leftarrow B \leftarrow C$ share the same conditional independence property as the common cause structure $A \leftarrow B \rightarrow C$, namely, the two terminal variables (A and C) in such structures are rendered independent once the value of the central variable B becomes known through either observation or intervention. In other words, we have that $(A \perp\!\!\!\perp C) \mid B$ in each of these three structures. It is also true that without observation or intervention on variable B, the terminal variables A and C will be dependent in probability within each of these three structures. Structures such as these which share the same conditional and marginal independence properties are termed *Markov equivalent* structures (c.f., Pearl, 1988; Lauritzen, 1999; Koller and Friedman, 2009). Furthermore, since all the structures within this Markov equivalence class share the same conditional independence property $(A \perp\!\!\!\perp C) \mid B$, we can label the equivalence class via this conditional independence property that is shared by all its members, i.e., we can use the label $\{A \perp\!\!\!\perp C \mid B\}$ to refer collectively to this entire equivalence class of directed structures.

From a marketing standpoint, the importance of identifying a structure's Markov equivalence class is that *each structure within the same Markov equivalence class will be indistinguishable based on observations alone*. In other words, no set of observations will be capable of singling out any particular member of a Markov equivalence class over any other member of that same Markov class (Lauritzen, 1999; Pearl, 2000).

As an example of this remarkable fact, consider the three members of the Markov equivalence class $\{A \perp\!\!\!\perp C \mid B\}$ which was discussed earlier. It can be demonstrated that any member of this Markov class can be transformed into any other member of this class through applications of probabilistic identities such as Bayes' theorem. For instance, suppose that we begin with the causal chain $A \rightarrow B \rightarrow C$ and we wish to demonstrate that it can be transformed into the common cause structure $A \leftarrow B \rightarrow C$. Essentially, this can be accomplished via reversal of the $A \rightarrow B$ link in the structure $A \rightarrow B \rightarrow C$ through Bayes' theorem, which can be calculated as shown in Equations 2a through 2c, below.

$$
\begin{aligned}
P(A,B,C) &= P(C|B) * P(B|A) * P(A) && \text{(2a)} \\
&= P(C|B) * \frac{P(A|B) \ * \ P(B)}{P(A)} * P(A) && \text{(2b)} \\
&= P(C|B) * P(A|B) * P(B) && \text{(2c)}
\end{aligned}
$$

Note that the resulting factored distribution that occurs in Equation 2c corresponds to the common cause structure $A \leftarrow B \rightarrow C$, and hence we have shown that the common cause structure can be derived from the causal chain structure. In fact, as shown in Figure 8, we can utilize such Bayes' theorem based transformations to convert any specific member of this Markov equivalence class into any other member of this same Markov class.

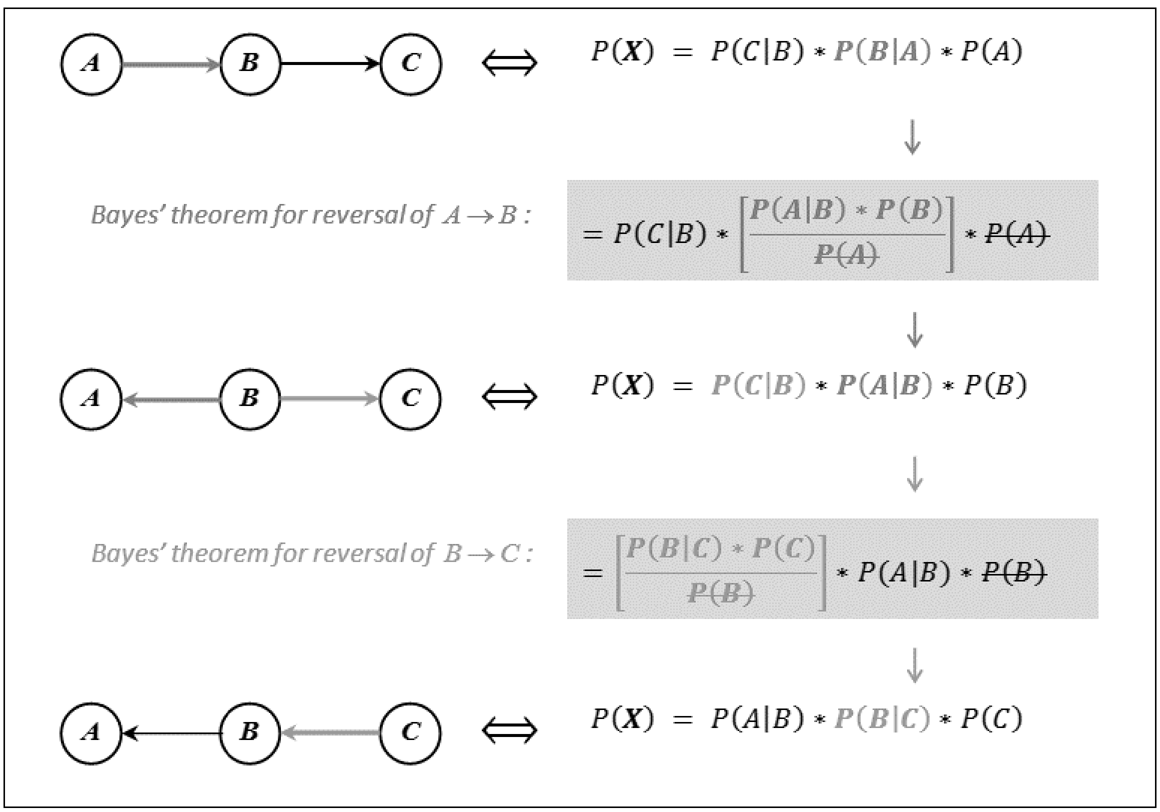


Figure 8. Equivalence of the members of the $\{A \perp\!\!\!\perp C \mid B\}$ Markov equivalence class

We can also examine these equivalences from a consumer's standpoint. For instance, consumers are likely to infer facts or properties about the relationships among the variables related to a product or brand based on the comparative rates at which these variables tend to occur or to co-occur within a set of observations (Hoch and Ha, 1986). Now consider the three members of the $\{A \perp C \mid B\}$ Markov equivalence class, namely the common-cause structure $(A \leftarrow B \rightarrow C)$, the causal chain $(A \rightarrow B \rightarrow C)$ and the diagnostic chain $(A \leftarrow B \leftarrow C)$. In each case, the exogenous variable (namely variable B for the common effect structure, variable A for the causal chain, and variable C for the diagnostic chain) will tend to occur at its own specific exogenously determined rate, which the consumer will typically be interested in monitoring as a primary goal of his or her information acquisition process regarding this particular product or brand. However, due to the directional relations among the variables in each of these structures, the rates of occurrence of the exogenous variable in each case will tend to get transmitted via the directed links within that structure to the variables which are probabilistically linked to those exogenous variables. This rate of transmission of probabilistic information will then result in specific co-occurrence rates for each of the linked pairs of variables.

For example, consider the common-cause structure $(A \leftarrow B \rightarrow C)$. Assuming that the rates of spontaneous occurrence for each of the three variables are minimally correlated, then the rates of *co-occurrence* of variable pairs {B,C} and {A,B} will be controlled by the causal mechanisms $B \rightarrow C$ and $B \rightarrow A$ respectively. Therefore, as long as these causal strengths are roughly comparable, the two variable pairs {B,C} and {A,B} will each occur at about the same rate as one another. Similarly, in the causal reasoning chain ($A \rightarrow B \rightarrow C$), variable A's occurrence will lead to that of variable B, which in turn will lead to the occurrence of variable C, and hence each of the variable pairs {B,C} and {A,B} will once again occur at roughly equal rates. Finally, in the diagnostic reasoning chain ($A \leftarrow B \leftarrow C$), the two variable pairs {B,C} and {A,B} will again occur with roughly equal rates, since the occurrence of variable C leads to that of B and A.[23] Lastly, in all three structures, due to the roughly equal co-occurrence rates of variable pairs {B,C} and {A,B}, we will also see a significant correlation among all three variables in each structure. Hence, just as each of these Markov equivalent structures can be transformed into one another through probabilistic equivalences, so too are these structures equally capable of supporting any particular set of observations. A consumer observing these variables' rates of co-occurrence would not be able to tell which of the Markov equivalent structures was actually the underlying mechanism which generated that set of product or brand-related observations.

On the other hand, one can demonstrate that none of the three causal structures in the $\{A \perp C \mid B\}$ Markov equivalence class can be transformed into a *common-effect* structure (such as the example involving the price of a stereo in Figure 5). For instance, attempting to convert the causal chain structure $A \rightarrow B \rightarrow C$ into the common-effect structure $A \rightarrow B \leftarrow C$ by reversing the $B \rightarrow C$ link would result in the following series of algebraic transformations:

[23] This argument is a slight modification of that given in Steyvers, et al. (2003).

$$P(A, B, C) = P(C|B) * P(B|A) * P(A) \tag{3a}$$

$$= \frac{P(B|C) \ * \ P(C)}{P(B)} * P(B|A) * P(A) \tag{3b}$$

The resulting expression in Equation 3b does not further simplify, and is obviously not equivalent to the structure of the common-effect model $\mathrm{A} \rightarrow \mathrm{B} \leftarrow \mathrm{C}$, which has a joint probability distribution with a factored form that is given by the expression $P(A, B, C) = P(B|A) * P(A) * P(B|C) * P(C)$. Therefore, the simple "reversal" of the $\mathrm{B} \rightarrow \mathrm{C}$ link in the causal chain $\mathrm{A} \rightarrow \mathrm{B} \rightarrow \mathrm{C}$ does *not* result in the common-effect structure $\mathrm{A} \rightarrow \mathrm{B} \leftarrow \mathrm{C}$. Rather, these two are fundamentally different causal structures belonging to different Markov equivalence classes; one cannot be transformed into the other through probabilistic equivalences.[24]

Evidently, there is more to link directionality than just drawing arrows between variables. There is an entire probabilistic structure "buried" within the structure of these links, and this probabilistic structure regulates when link reversal is possible and when it is not. As shown, some link directions can be reversed without altering the conditional independence properties of the overall network (such as in the successful conversion of the causal chain $\mathrm{A} \rightarrow \mathrm{B} \rightarrow \mathrm{C}$ into the common-cause structure $\mathrm{A} \leftarrow \mathrm{B} \rightarrow \mathrm{C}$), while

---

[24] Furthermore, from a consumer's standpoint, if all three variables (A, B, and C) occur at independent rates, then in the case of a common-effect structure $\mathrm{A} \rightarrow \mathrm{B} \leftarrow \mathrm{C}$, the variable pairs {A,B} and {C,B} will also occur at differing rates (as dictated by the independent rates of occurrence of the exogenous variables A and C respectively). Such differences in the rates of co-occurrence of the variable pairs {A,B} and {C,B} can therefore alert the consumer that there are likely to be two independent causal or directional mechanisms that can lead to variable B. This differs significantly from the conclusions that can be drawn from any of the three members of the $\{A \perp\!\!\!\perp C \mid B\}$ Markov equivalence class discussed earlier, in which the variable pairs {A,B} and {C,B} will typically occur at equal rates within any of the three members of that equivalence class of directional structures.

other links cannot be reversed in this manner (such as in the failed attempt to convert the causal chain $A \to B \to C$ into the common-effect structure $A \to B \leftarrow C$). Link directionality is only "fungible" up to a certain point, but not beyond it. Hence the introduction of directionality into an otherwise undirected brand concept map requires more than just an introduction of directional arrows where there formerly were none. Rather, there is an entirely different structural semantics in a directed network, namely a semantics dictated by the conditional independence properties of the data and the Markov equivalence properties of the resulting directed structures.

2. Bayesian Networks and Brand Concepts

2.1 Observation, Intervention, and Markov Equivalence

It is clear from our earlier discussion that consumers routinely view all manner of marketing phenomena as having directional properties. However, as we have seen in the previous section, one cannot simply 'introduce directions' into an otherwise undirected brand concept network. Rather, such directions interact with one another in a manner that is controlled by the conditional independence properties inherent in the data and the Markov equivalence classes to which the various directional structures and substructures within that network belong. Hence it is critical to examine how consumers may interact with and utilize the directional structure of a brand concept network in order to enhance their understanding of the possible relations among a brand's associations.

For example, as we have seen, the various directional structures that can exist within a directional network can be grouped into Markov equivalence classes, each of which contains structures that cannot be distinguished from one another based on sets of observations of the variables involved. However, consumers depend on understanding causal structures in order to be able to determine which variables to manipulate or select in order to create conditions which favor specific desired outcomes (Hoch and Deighton, 1989). Furthermore, consumers also depend on understanding causal relations among brand-related variables in order to attribute credit or blame for either good or poor product or brand performance (Folkes, 1988; Weiner, 2000). Hence, the fact that directional structures within a brand concept network cannot be distinguished beyond the level of a group of Markov equivalent structures might be regarded as problematic by a consumer who wishes to determine which variables or variable levels to select when assembling a consideration set for a particular purchase, or when making an attributional determination concerning a particular product or brand.

If the various directional structures within a Markov equivalence class all represented the same set of causal or predictive relationships, then the observational indistinguishability of the structures within such an equivalence class would not pose a problem. However, Markov equivalent structures typically do not indicate the same set of causal or directed relationships. For example, both the causal chain $\mathrm{A} \rightarrow \mathrm{B} \rightarrow \mathrm{C}$ and the common effect $\mathrm{A} \rightarrow \mathrm{B} \leftarrow \mathrm{C}$ are members of the same Markov equivalence class (namely the class $\{\mathrm{A} \perp \mathrm{C} \mid \mathrm{B}\}$), yet each of these structures will clearly indicate a different set of causal relationships between the three brand associations A, B, and C.

Interestingly however, even though the structures within the same Markov class are indistinguishable through observations alone, consumers *can* actually “peer into”

Markov equivalence classes and begin to differentiate among the various constituent causal structures by combining observations of these variables with *interventions* on one or more of those variables. For example, consider the causal chain $A \rightarrow B \rightarrow C$ and the common-cause structure $A \leftarrow B \rightarrow C$. As argued previously, if we passively observe the patterns of co-occurrence of the variables within these structures, the source of variation will reside in the exogenous variable in each case, and since each of these exogenous variables is either a direct or indirect cause of the other two variables in its respective causal structure, then under observational conditions we would simply expect to see all three variables in each structure either become jointly activated or remain jointly inactivated. However, if a consumer *intervenes* to set the level of one or more variables rather than letting an exogenous cause generate the chosen activation levels, then that consumer should be able to discern one causal structure from another, *even in cases where the respective causal structures belong to the same Markov equivalence class, and hence are observationally indistinguishable from one another*.

As an example, consider a consumer evaluating a set of products, each of which has some probability of possessing three binary attributes A, B, and C. As we have already argued, no amount of observation of different product exemplars will be sufficient to guarantee that this consumer can distinguish among the three potential generative causal models (causal chain, diagnostic chain, and common cause) belonging to the $\{ A \perp C \mid B \}$ Markov equivalence class. However, suppose that our consumer *intervenes* upon the causal structure by, for example, considering only those products which are known to possess attribute B. In such a case, rather than allowing exogenous conditions to determine whether or not attribute B is present, the consumer himself has set the 'level' (or 'value') of attribute B to 'on'.

Now consider the effect that this intervention by our consumer has upon the three different observationally indistinguishable causal structures belonging to the $\{A \perp C \mid B\}$ Markov equivalence class. Since the level of attribute B is now established by the choice of the consumer, all other causal influences on the value of this attribute are effectively eliminated from consideration. For example, if the true underlying structure were actually a causal chain ($A \rightarrow B \rightarrow C$), then the causal influence $A \rightarrow B$ of attribute A upon the value of attribute B is eliminated by this consumer's intervention to 'fix' the level of attribute B to 'on' (e.g., through refinement of his consideration set to just those products known to possess attribute B). Similar considerations also apply as well when considering the effect of such an intervention upon the other members of this equivalence class of otherwise observationally indistinguishable causal (or 'generative') structures that potentially underlie the observed set of interdependencies among product attributes A, B, and C.

One can, in fact, determine that our consumer's intervention upon this set of observationally indistinguishable generative structures serves to split this single Markov equivalence class into three separate equivalence classes, each of which is now observationally distinguishable from the others (c.f., Steyvers et al., 2003). The main principle underlying such a transformation is that by fixing the level of an attribute (such as attribute B in this case) to a specific level, the consumer is effectively *eliminating* the effects of all other variables that are graphical *parents* of the intervened-on variable, but is *maintaining* the causal effect of the intervened-on variable upon all of its graphical *children* (e.g., Pearl, 2000). For example, in the case of the consumer who fixes the value of variable B by considering only those products known to possess attribute B, the single Markov equivalence class { A ⊥ C | B } of underlying generative causal structures is split into three different causal structures, as shown in Figure 9, below.[25]

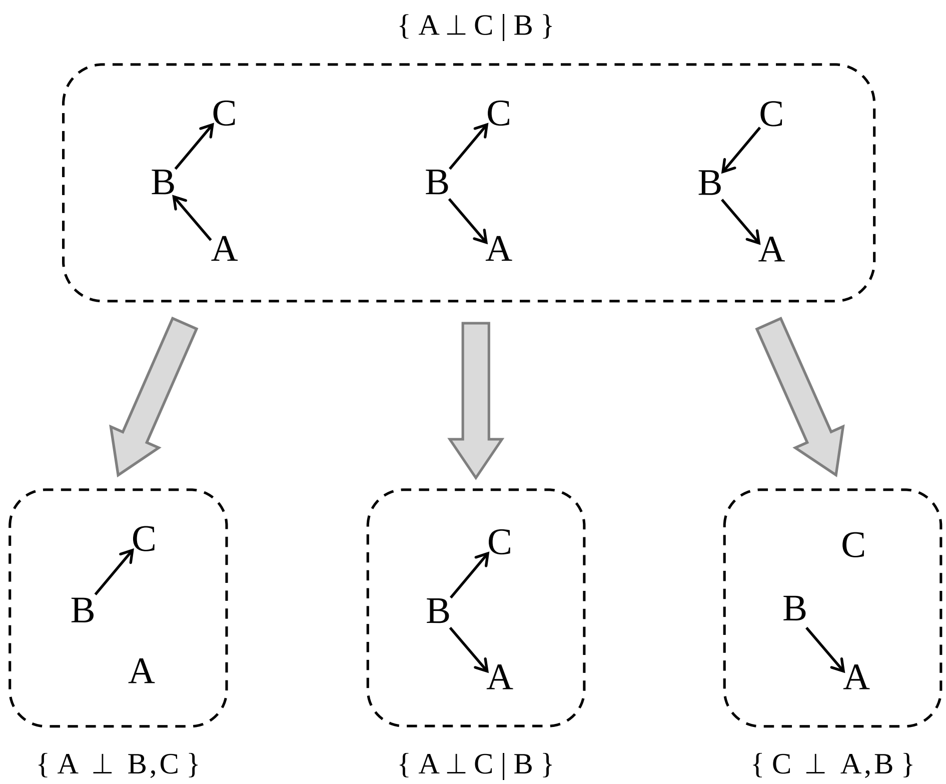


Figure 9. Splitting the equivalence class {A ⊥ C | B} through intervention on B

[25] An alternate version of this diagram can be found in Steyvers, et al. (2003).

As Figure 9 illustrates, a consumer who intervenes to 'fix' the value of variable B (for instance by limiting his consideration set to just those products containing attribute B) will observe one of three different resulting patterns of dependence. In the leftmost case shown in Figure 9, we are assuming that the member of the Markov equivalence class $\{ A \perp C \mid B \}$ which was responsible for the observed correlation among attributes A, B, and C is the causal chain structure $A \rightarrow B \rightarrow C$. Since the consumer has selected only those product exemplars which possess attribute B, the presence or absence of variable A no longer plays a causal role in influencing whether or not these exemplars will possess attribute B (since they already do possess the attribute). Therefore, such a consumer will observe that attributes B and C tend to occur together, but that attribute A occurs with its original exogenously determined frequency which is unrelated to the occurrence of attributes B and C. This would be akin to a consumer who evaluates a particular category of products and notices a correlation between the perceived price level (high/low), the perceived quality (high/low), and the perceived channel exclusivity (high/low) for these products. Should such a consumer find that upon considering only those products perceived to be of high quality, that price is no longer related to channel exclusivity, he or she would likely conclude that it was quality all along that had driven channel choice, rather than the price itself driving channel choice directly.

In a similar manner, should the underlying generative structure in Figure 9 be the right-hand causal structure $C \rightarrow B \rightarrow A$ (the so-called diagnostic reasoning chain), intervention by the consumer to fix the level of attribute B would result in attribute B co-occurring with attribute A, but attribute C becoming unrelated to the frequency of occurrence of either of these two other variables. On the other hand, should the underlying causal model actually be the common-cause structure shown in the middle of Figure 9, intervention to select only products possessing attribute B would not make any change to the observed relationships among these three variables. The occurrence of variable B will still drive the occurrence of both attribute A and attribute C, and since attribute B is the exogenous variable in this causal system, all three attributes will tend to occur together.

In terms of the previously discussed consumer who is evaluating a particular class of products and who believes there to be a correlation between the price level, the quality, and the channel exclusivity for these products, his or her intervention upon this product category through considering only those products containing attribute B (high quality) would, in the case of the common-cause structure shown in the middle of Figure 9, still result in all three attributes (price, quality, exclusivity) occurring together. Since restricting the value of the ‘quality’ attribute to ‘on’ has not disturbed the rates of co-occurrence of the different variable pairs in this system, such a consumer would likely conclude that the intervened-upon variable ‘product quality’ was a common cause of the other observed variables rather than being an effect of either one of them. After all, if a variable (such as variable B in the causal structures in Figure 9) were an effect of one or more other variables (such as variables A or C for instance), then fixing the level of variable B should indeed serve to break the dependence of variable B upon the values of those supposed causes.

Interestingly, consumers do not necessarily have to actually *perform* interventions such as this, which serve to 'fix' the values of particular variables or attributes at certain levels, in order to reason about their likely effects. Rather, such reasoning on the part of consumers can be *counterfactual* in nature (McGill, 2000). For instance, a consumer in the previous example would not necessarily need to purchase items for which the value of attribute 'B' (high quality) is set to 'on' in order to reason about the likely effect of such a restriction. Rather, he or she can merely *consider* what would happen if such a restriction were to be made (Krishnamurthy and Sivaraman, 2002). In fact, consumer theorizing and fantasizing about the potential effects of various consumption choices is an important determinant of consumer preference and satisfaction (Holbrook and Hirschman, 1982), and more generally, counterfactual thinking is known to be a core component of normal cognitive and social functioning (Summerville and Roese, 2008).

Counterfactual reasoning can be considered as a form of "guided thought experiment" in which events that are potential causes of other events are mentally negated (Kahnemann & Tversky, 1982; Klayman & Ha, 1987)) or mentally enacted or 'set' to specific values (Dunning & Parpal, 1989; Markman et al., 2007) and the effect of this intervention on other causally related variables is assessed. We can utilize such counterfactual manipulations to mentally "test" which possible causal structures from among a group of observationally equivalent models is the most plausible based on the data or our general knowledge and beliefs about the causal domain (Woodward, 2003).

The notion of testing which directional structure from within a class of observationally equivalent structures is likely responsible for generating a set of observations quite closely parallels the differences between observational findings and experimental findings within the process of normal scientific investigation (Woodward, 2007). Specifically, within an observational study, one can only ascertain that certain values of one event or variable tend to occur within the same datum as certain values of another event or variable, i.e., a correlational finding. In order to make the claim that manipulation of an independent variable will affect a specific dependent variable, one must intervene on that structure (Babbie, 1998).

More specifically, in a controlled scientific intervention, one changes the independent variable from value *x* to value *y*, but all other relationships are kept intact. On the other hand, in a correlational study, one does not know what other changes are occurring in the system, and therefore one does not have license to make the *counterfactual assertion* that had the independent variable *not* changed from value *x* to value *y*, the dependent variable would *not* have changed from *a* to *b*. In other words, whereas a correlational finding allows us to say what we observed happening in an undisturbed system, a causal finding allows us to make *counterfactual* claims about what would and would not happen among several different possible scenarios that differ from one another in specific and known ways (Hunt 1991: 112). As Roese & Olson (1996) summarize the matter, "asserting that the addition or deletion of antecedent X 'undoes' outcome Y leads to the causal attribution that X caused Y."

Within the consumer behavior literature, Hoch and Ha (1986) point out that consumers treat informational claims in the marketplace as tentative hypotheses about specific products, brands, and services, and then make purchase and usage choices which act as specific tests of these hypotheses. Hoch and Deighton (1989) add that consumers actively construct hypotheses as sets of " 'if p then q' condition-action rules" created through the processes of abduction and generalization, and then inductively strengthened through subsequent product experience. Eisenstein and Hutchinson (2006) describe such experiential market-based learning as "action-based" learning, which is the result of "repeatedly making decisions about concrete actions and then observing the outcomes." Furthermore, as argued by Lagnado et al. (2007), the fact that interventions are freely-chosen human actions allows individuals to create a quasi-experimental conditions since such freely-chosen human actions are essentially an intuitive means for individuals to remove or reduce confounds. Even more generally, as Lagnado and Sloman (2004) state,

> We are continually conducting informal experiments of our own to learn about the world around us. We remove items from our diet to see what is making us unhealthy or overweight, we tinker with new software programs to see what does what, we experiment with different ingredients in search of a tasty meal. (Lagnado and Sloman, 2004, p. 856)

Hence individual decision makers also manipulate their environment in an attempt to determine which choices lead to which outcomes, and thus which member of a set of observationally indistinguishable causal structures is the most likely generative mechanism that is operative within a product or brand domain. We are all informal scientists in this respect.

Furthermore, since the determination of causal structure is so paramount for consumers' understanding and decision-making processes (Folkes, 1988; Weiner, 2000), it stands to reason that any structures which can regulate or determine when consumers will require the use of interventional reasoning in order to ascertain needed facts about a market will be of paramount importance. Clearly, the notion of when various market configurations are observationally (or correlationally) indistinguishable from one another is such a critical construct, since this will determine when consumers are most likely to intervene and make choices in a market. As we have seen, observationally or correlationally indistinguishable structures are known as Markov equivalent structures, and hence the Markov equivalence construct provides an essential framework for understanding both the rationale and intended content of consumers' marketplace interventions.

2.2 Bayesian Networks and the Screening-Off Condition

Our main goal in this thesis is to determine an appropriate directional semantics for understanding consumers' brand constructs. Hence it will be quite useful to place some fairly commonsense limits on the types of directed structures which we will utilize to model such marketing-related phenomena. Specifically, in this thesis we will limit our modeling focus to the most widely-used category of directed structures, namely the category of *Bayesian networks*.

As defined by Pearl (1988), a Bayesian network is a directional model in which each node represents a variable from the domain being modeled, and the links between nodes represent directional influences between the variables connected by that link.[26] In addition, a Bayesian network must possess the following three regularity properties:

(1) The graphical structure must not contain any directed 'cycles' (i.e., it contains no 'infinite loops'). This condition is meant to insure that there is a clear flow of information among the variables within the domain, and information never gets 'trapped' inside an infinite regress or continual 'loop'. Directed graphical structures with no such infinitely repeating loops are called Directed Acyclic Graphs (or DAG's for short). Hence this condition can be summarized by saying that a Bayesian network must have a DAG structure.

(2) Within a Bayesian network graphical structure, the immediate graphical 'parents' of a variable must be capable of rendering that variable independent of all other variables besides its effects (or graphical "children"). This condition is generally known as the *causal Markov condition* (Pearl, 1988), or the *screening-off condition* (Sloman, 2005). Hence, under this condition, once the state of a variable's parents is known, the only remaining way to affect the state of that variable is through *diagnostic reasoning* 'backwards' from its effects, or graphical 'children'. Therefore, under this condition, we are able to avoid having to reason through lengthy chains of indirect causal influences in order to determine the value of a particular variable (Sloman, 2005).

[26] Some authors also require that the network parameters of the form $P(X_i \mid pa(X_i))$ also be fully specified. In other words, there is both a qualitative component of a Bayesian network which consists of that network's overall topology and directional structure, and a quantitative component which consists of the resulting conditional distributions at each node of that network (e.g., Pearl, 1988).

(3) All probabilistic independencies that exist among the variables in the causal graph must be the result of the structure of the connections among the variables in the graph, and not due to any unusual (or "unstable") coincidences of the various marginal and conditional probabilities involved. Formally, this condition is known as *causal faithfulness*, and it is meant to rule out "accidental" independencies, such as would be the case, for instance, if two different causal pathways with precisely opposite weights connected two variables A and B within a causal model, so that the effects of the two different causal pathways exactly cancel each other out. Such an apparent independence between A and B would merely be the result of a coincidental (and unstable) set of parameter values, and hence would not be the result of an application of the causal Markov condition to the structure of the corresponding causal graph. Such coincidental and unstable arrangements are precisely what the faithfulness condition guards against. In this sense, the faithfulness condition simply reflects people's natural desire to base their perceptions of a domain upon stable generative mechanisms rather than upon rare numerical coincidences (Sloman, 2005).

From a consumer modeling standpoint, the most important of these three properties is the causal Markov condition, since this property delineates precisely why a Bayesian network forms a much more cognitively efficient model of human reasoning within probabilistic domains as compared to an undirected graphical model of that same domain. For instance, consider a set of variables $X = \{V_1, V_2, . . . , V_n\}$, where n may be a fairly large number if this domain is reasonably large. In general the joint probability distribution P(X) of this set of variables would need to be expressed using the chain rule of probability as:

$$P(X) = P(V_n \mid V_1, V_2, \ldots, V_{n-1})\, P(V_{n-1} \mid V_1, V_2, \ldots, V_{n-2}) \cdots P(V_2 \mid V_1)\, P(V_1) \qquad (4)$$

Even in the 'best case scenario' in which each variable in this domain is binary, such a computation would still require an immense number of probabilities to be estimated from the data, and it is difficult to imagine a typical consumer being able to cognitively manage such a computation at all, much less within the time frame allotted to typical shopping decisions involving domain variables such as these. On the other hand, if these variables exist within a Bayesian network representation of the domain, then due to the causal Markov condition, the overall probability distribution P(X) can simply be expressed as $P(X) = \prod_{i=1}^{n} P(V_i \mid pa(V_i))$, where $pa(V_i)$ denotes the graphical 'parents' (i.e., the immediate graphical ancestors) of variable V_i within the Bayesian network representation (Pearl, 1988; Lauritzen, 1999; Koller and Friedman, 2009).

For example, consider the Bayesian network example given in Figure 10. [27] Even in the 'best' case, in which all fourteen variables are binary, computation of the full joint probability distribution in this domain would require over 16,000 probabilities to be estimated (and the problem gets exponentially worse if one or more of the variables involved contains more than just two possible values). In fact, by the chain rule of

[27] Note that even though there is a 'loop' structure in this network configuration, the directions within the loop do not allow one to 'cycle around' forever. Hence this loop structure is not a directed cycle, and therefore this structure is a directed acyclic graph (or DAG).

probability, the full expression for the probability distribution P(X) for this entire joint density would take the form:

$$P(M|A,B,C,D,E,F,G,H,I,J,K,L)\ P(L|A,B,C,D,E,F,G,H,I,J,K) \cdots P(B|A)\ P(A) \qquad (5)$$

However, by employing the causal Markov condition within this Bayesian network representation, we can reduce this enormous multivariate density containing over 16,000 parameters to a simple product of very small local distributions, each of which typically involves just a few variables and a correspondingly small number of parameters. For example, since variable M has only one graphical parent, we do not need to condition its distribution on all 13 remaining variables (as is done in the first factor in Equation 5). Rather, we can simply condition variable M on its single graphical parent, variable K. Similar reductions are possible throughout this Bayesian network. In fact, by using the causal Markov property, one can reduce the massive probability density expression in Equation 5 to the much simpler expression given in Equation 6 (in which we have omitted the symbol "P" in front of each argument for compactness sake).

$$(A)\ (B|A)\ (C|B)\ (D|C)\ (E|C)\ (F|B,D)\ (G|F)\ (H|G)\ (I|H)\ (J)\ (K|G,J)\ (L|J)\ (M|K)\ (N|I) \qquad (6)$$

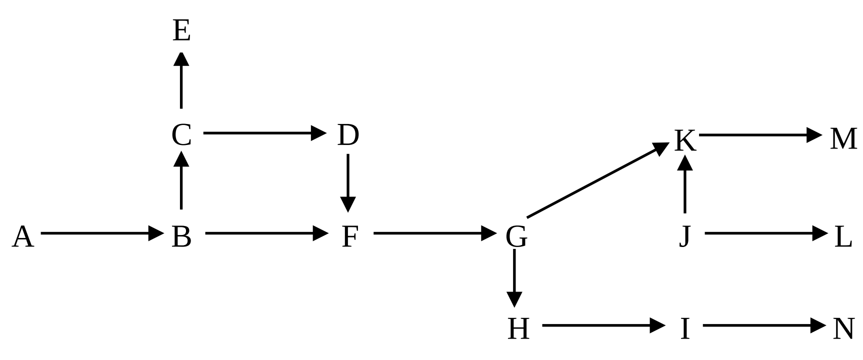


Figure 10. Bayesian network example with 14 variables

Note that the reduction of the full density expression of Equation 5 to the much simpler representation of Equation 6 is possible because the Bayesian network representation replaces each global factor in the full expression by a much more compact local factor. This is achieved because the Markov condition in a Bayesian network enables the parent nodes of a variable to render that child node independent of the effects of any of its nondescendant nodes (including additional 'ancestor' nodes which lead to that child node's parents), and hence the only knowledge that is necessary to specify the distribution of a node within a Bayesian network is the conditional distribution of that node given the various states of its direct parent nodes, along with the state of those parent nodes. Complex, global factors become reduced to simple, local factors defined over far fewer nodes. For example, whereas a direct determination of the full joint distribution over the 14 variables shown in Figure 10 would require specification of over 16,000 parameters (even in the 'best case' scenario of completely binary variables), the causal Markov condition applied to the corresponding Bayesian network allows us to reduce this number to merely 30 parameters required to fully specify the joint distribution.

Now consider the reasoning task required of a consumer scanning the myriad products available in a typical supermarket trip, for instance. As commented before, such a consumer routinely encounters tens of thousands of products in a typical shopping exercise such as this (Broniarczyk, 2006), and hence an undirected and/or minimally structured cognitive framework that represents the various brands, products, and related characteristics involved in a shopping scenario such as this would almost certainly be beyond the ability of any consumer to fully comprehend. Hence a natural question to ask is whether such a consumer could potentially employ a similar complexity-reduction strategy to that utilized within a Bayesian network in order to effectively reduce this incomprehensibly large set of product and attribute combinations to a manageable and much more cognitively tractable structure.

Clearly, a consumer is unlikely to derive and parameterize a Bayesian network. However, consider the fact that the operative principle which enables a Bayesian network to drastically simplify the representation of a domain is that of conditional independence. For instance, when a Bayesian network representation reduces a complex global set of factors to a much simpler set of local factors computed over a far smaller set of nodes, such a reduction is afforded by the ability of the graphical parents of a node to render their child node conditionally independent of all other nodes which are not descendants of that child node in question.

Interestingly, human reasoners do routinely utilize an intuitive cognitive analog of the conditional independence relation: namely the principle of conditional irrelevance. This principle allows a reasoner to decide that knowledge of the state of one or more variables is sufficient to allow other variables to be effectively ignored (or severely discounted in importance). For instance, consider the following example described by Pearl (1988):

> A person who is reluctant to estimate the probability of being burglarized the next day or of having a nuclear war within five years can nevertheless state with ease whether the two events are dependent, namely, whether knowing the truth of one

proposition will alter the belief in the other. Likewise, people tend to judge the three-place relationship of conditional dependency (i.e., X influences Y, given Z) with clarity, conviction, and consistency. For example, knowing the time of the last pickup from a bus stop is undeniably relevant for assessing how long we must wait for the next bus. However, once we learn the whereabouts of the next bus, the previous knowledge no longer provides useful information.

(Pearl, 1988, p. 79)

In this example, the time of the last pickup and the time of the next pickup are certainly correlated variables: knowing that the last pickup was recent reduces the probability that the next bus will arrive soon. However, once we learn the location of the next bus, the variable "time of last pickup" no longer influences the variable "time of next pickup". Stated another way, these two variables have become conditionally independent given this additional piece of knowledge. Statistically, this conditional irrelevance relationship is: $next\ pickup \perp\!\!\!\perp previous\ pickup \mid location\ of\ next\ bus$.

Similarly, in a marketing context, a consumer might believe that the price of a product and the level of quality of that product are correlated, but become independent once we know that the product comes from a specific brand. For instance, such a consumer might think that generally, the more expensive an automobile is, the higher quality it is likely to have. However, this consumer might also believe that once he or she knows the car is a BMW, then price no longer determines quality for this car because he or she feels that all BMW automobiles are of high quality, and within the BMW brand, higher price just buys more features but not additional quality. Hence, the brand concept map for such a consumer should contain the conditional independence property: $price \perp\!\!\!\perp quality \mid BMW.$

At an even more basic level, consider the simple act of ignoring certain product or brand attributes altogether. For example, some types of consumers tend to ignore packaging differences among products. However, rather than being a unary relation of the form "packaging is irrelevant", this phenomenon may actually be a trinary relation of the form "given the product's ingredients, the packaging becomes irrelevant to the expected performance of the product." [28] In more generality, if some product or brand feature is considered irrelevant by a consumer, it must be irrelevant *to* something, i.e., irrelevant to some particular desired state that the consumer is seeking with regards to that product. However, even a feature that is considered irrelevant by the consumer is still associated with that product or brand in some manner (since otherwise it would not be considered a brand or product feature, irrelevant or otherwise). Hence such a feature could, in theory, have *some* relevance to that desired state or end goal. It is merely that some *other* product or brand feature provides enough of a clue to that desired state or end goal to render any additional clues irrelevant in the mind of the consumer.[29, 30]

[28] After all, if the ingredients were completely unknown, then aspects of the packaging might become a secondary source of clues about the product's intended positioning or functioning.

[29] In fact, this conditional irrelevance construct (i.e., aspect x is irrelevant to goal y in the presence of another aspect z) may be so innate or so deeply embedded within the means by which we cognitively deal with large consumption domains that we typically take it for granted that a feature that is considered irrelevant is relegated to such irrelevance by the presence of other features which are more indicative (either predictively or diagnostically) of the desired goal being sought with such a product or brand.

[30] It could be that some product or brand features are truly irrelevant *ab initio*. However, we maintain that a significant number of product or brand features may be irrelevant to a desired state or end goal because the presence of some other product or brand feature renders them irrelevant. Hence, the delineation of how the conditional irrelevance construct interacts with other aspects of consumers' brand concepts may offer a valuable contribution to the understanding of the motivations and decision-making apparatus that drive consumption decisions.

In conclusion, what we have argued for here is the notion that the same 'screening off' process that allows a Bayesian network model of a consumption domain to reduce a set of complex global factors to a much simpler set of small, local factors, thereby rendering the joint density of those marketplace variables much more readily computable is precisely the same construct that allows consumers to 'screen off' many product or brand characteristics, thereby rendering a massively complex consumption domain cognitively tractable. Thus, we are arguing that the relation which a statistician terms conditional independence is also conceived of by consumers in a more innate fashion as simply 'conditional irrelevance', i.e., the notion that knowing some brand or product characteristics renders many other such characteristics irrelevant to the main goals being sought through that product's consideration or purchase.

2.3 Variable Separation Within Directed Structures

As we have seen in the preceding sections, the information semantics within a directed network are quite different than they are within an undirected network. For instance, the directions of the links in a Bayesian network representation of a domain entail specific conditional and marginal independence properties among those domain variables. Since the property of conditional independence is a three-place relation (i.e., X is independent of Y given Z), these conditional and marginal independence properties are graphically encoded within several basic three-variable directed structures that exist within a Bayesian network, namely the causal chain ($A \rightarrow B \rightarrow C$), the diagnostic chain ($A \leftarrow B \leftarrow C$), the common-cause structure ($A \leftarrow B \rightarrow C$), and the common-effect (or 'collider') structure ($A \rightarrow B \leftarrow C$). However, once we assemble these more basic structures into a larger-scale Bayesian network model, the various conditional and marginal independence properties represented by these basic structures can interact in more complex ways, making it difficult to judge specifically when one variable or set of variables can probabilistically "shield" one group of product or brand associations from the effects of the other such associations. In order to determine the resulting independence properties of the resulting structure, there must be a single criterion which can determine when two variables (or sets of variables) anywhere in the network are probabilistically separated from one another.[31] Such a criterion is known as the *directed separation principle* (or "d-separation" for short), e.g., Pearl (1988; 1990).

To understand the d-separation principle, recall that for a chain connection (either a causal chain or a diagnostic chain), information propagation becomes blocked if we observe (or set) the value of the central variable in the chain. Also recall that for a common cause structure ($A \leftarrow B \rightarrow C$), information propagation is also 'blocked' by knowledge of the value of the central variable in the connection. However, within a convergent connection (i.e., a common effect structure, or 'unshielded collider', such as $A \rightarrow B \leftarrow C$), information propagation is *facilitated* by knowledge of the value or distribution of the central variable, but is blocked when the value or distribution of the central variable is not known. Based on these principles, we can elucidate a global criterion which can determine when two variables or sets of variables anywhere in the

[31] Note that if two variables (or sets of variables) are probabilistically separated from one another, then they are conditionally independent given the values of the variables which separate them. Thus, this separation criterion can also be referred to by the conditional independence properties that it encompasses.

network are probabilistically separated (and hence are conditionally independent). This directed separation (or 'd-separation') principle can be stated as follows (Korb and Nicholson, 2004):

<u>d-separation principle</u>[32] :

Given variables X and Y, along with a set of variables Z disjoint from both X and Y, the variables X and Y are *d-separated* given Z if and only if all paths Φ between X and Y are "cut" by one or more of the following graph-theoretic conditions:

1. Φ contains a chain $A \rightarrow B \rightarrow C$ or $A \leftarrow B \leftarrow C$ such that $B \in Z$.
2. Φ contains a common-cause connection $A \leftarrow B \rightarrow C$ such that $B \in Z$.
3. Φ contains an unshielded collider (or 'immorality') $A \rightarrow B \leftarrow C$ such that neither B nor any of B's descendants is in Z.

Note that these three conditions are essentially just formalizations of the separation properties which we have previously discussed for the four different types of directed triples that can exist within a Bayesian network. For example, Conditions (1) and (2) merely state that if a path between variables X and Y contains a subgraph that is isomorphic to either a causal chain, a diagnostic chain, or a common cause structure in which the central variable belongs to the putative separating set Z, then that separating set Z is capable of blocking information transmission along this path. Similarly, Condition (3) states that if a path between variables X and Y contains a subgraph that is isomorphic to an

[32] There are several equivalent statements of the d-separation criterion, as well as an alternative means of detecting conditional probabilistic independence, termed the *directed global Markov condition*. A more complete discussion of these various additional methods for determining when information propagation between two groups of variables is blocked can be found in the Appendix.

unshielded common effect structure, then the lack of an observation of the value of the central variable in this structure or any of its descendants can block information transmission along this path.

What is truly interesting about the d-separation principle is that by formalizing these earlier observations about the separation properties engendered by each of the different three-variable directed substructures, this principle allows one to formalize the relation between the topological properties of the graph and the properties of those probability distributions which describe the conditional and marginal probabilities at each node of that graph. Specifically, it can be shown (e.g., Verma and Pearl, 1988; Pearl, Geiger, and Verma, 1989; Geiger and Pearl, 1990; Pearl, 2000) that if a probability distribution over the variables in a particular Bayesian network generates conditional and marginal probability values at each node which satisfy the causal Markov condition with respect to that network, then the relation between conditional independence and the d-separation property within that network is both sound and complete: every d-separation in the network will represent a true conditional independency within that probability distribution, and every conditional independency in that distribution will be identified by the d-separation principle as applied to that network.

Another interesting fact related to directed separation within Bayesian network models is that due to the probabilistic differences between directed and undirected structures, d-separation and 'regular' (i.e., 'undirected') graphical separation do not always coincide. For instance, looking at the sample Bayesian network of Figure 10, it turns out that despite "appearing" like a separator of sorts, variable K does not actually d-separate variable G from variable J, since variable K occurs in an unshielded collider (or common effect) configuration, and hence knowledge of variable K will actually facilitate information propagation between variable G and variable J. On the other hand, variable K does d-separate variable G and variable M, since the single path from G to M (namely the pathway $G \rightarrow K \rightarrow M$) is a causal chain, and as such, would be blocked by knowledge of the value of variable K. Additionally, one can say that variables G and J are unconditionally d-separated (i.e., d-separated given $\emptyset$) since without knowledge of the value of the collider at variable K, there would be no way for G and J to probabilistically communicate.

Within a marketing domain, such considerations can be quite relevant because they provide practical guidelines for dealing with the introduction of directionality into otherwise undirected models. For instance, consider a situation in which a firm's product or brand has a set of established associations, and there is an additional influence on several of those variables (e.g., from a competing firm's actions, etc.) which the firm wishes to curtail. If this collection of brand associations is considered as an undirected network (such as would be the case via utilizing a modeling procedure such as the BCM) one might end up with a model such as that shown in the left panel of Figure 11. Within this model, let us assume that brand association B is a strategically central association that the firm wishes to protect, but which is not easily manipulable. Based on this undirected model, the firm's strategy seems simple: by fixing the values of attributes A and C, one would presumably block the ability for information from association E to enter this reasoning chain and influence variable B.

However, if the underlying model is a directed one (in which reasoning in one direction is predictive and reasoning in the opposite direction is diagnostic in nature), then the situation would actually be more appropriately modeled by the diagram shown in the right panel of Figure 11. In this case, fixing the values of attributes A and C would not serve to d-separate the strategically critical association B from the new association E. In fact, if the firm responded to this threat by fixing the values of attributes A and C, they would actually be creating a situation where attribute E and attribute B can become probabilistically related through the collider structure at attribute C. In a competitive situation such as this, the firm's best strategy would be to fix the value of just brand association A, and to try to make the value of brand association C (as well as any of its descendants, such as association D, etc.) as vague as possible in the mind of the consumer.

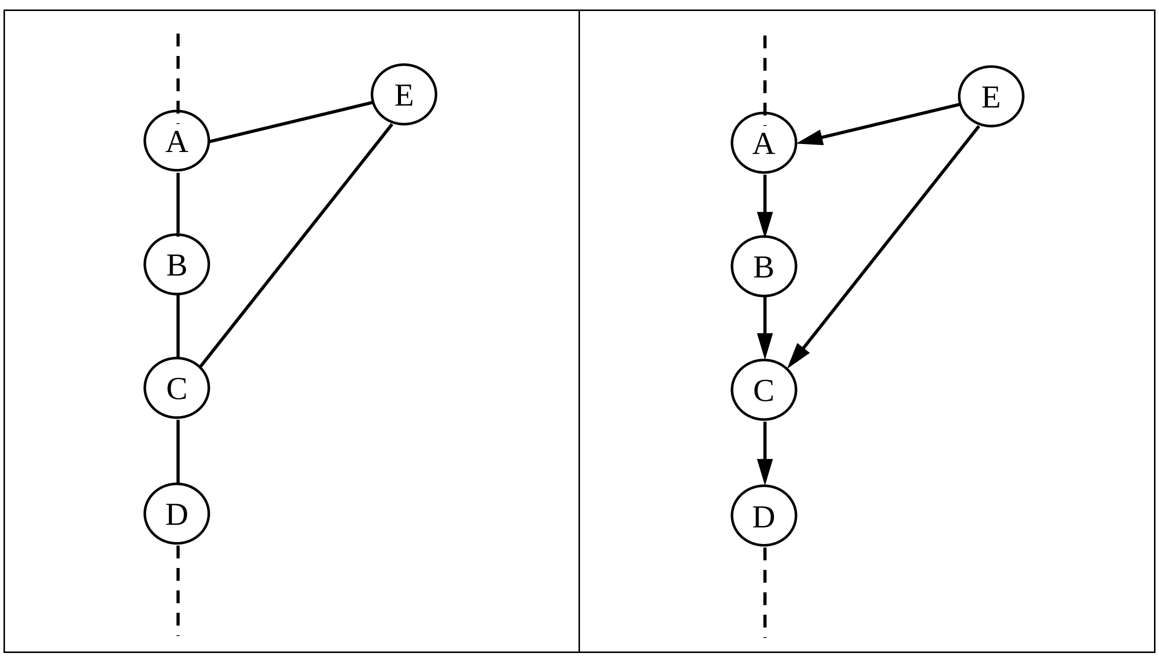


Figure 11. Directed and undirected models of a brand association scenario

2.4 Markov Equivalence and Causal Variation

Now that we have introduced the definition of a Bayesian network as a specific form of directed network that possesses certain regularity properties, and we have examined the associated constructs of directed separation between variables and the screening-off (or conditional irrelevance) property, we can apply these concepts to the elucidation of some truly novel and quite surprising marketing-related ramifications of the Markov (i.e., observational) equivalence of brand concept structures.

To begin with, there is a surprising set of criteria for determining precisely when two or more directed structures are Markov equivalent (i.e., when the structures encode the same set of conditional dependencies and independencies). To describe these criteria, we make the following two definitions: first, we define the *skeleton* of a Bayesian network to be the undirected structure that represents the link topology, with all notions of directionality removed. (In other words, the *skeleton* of a directed network structure is the undirected analog of that directed structure.) Secondly, we must single out one particular form of directed link configuration: a common-effect structure in which the two "parent" variables (or "causes") of the common effect are not themselves directly connected to one another. Such a configuration is often called an *immorality* within the Bayes net literature (since within such a structure, the two "parents" of the common effect, or "child", are "unmarried").[33]

[33] Such a structure is sometimes also given the more prosaic name "*uncovered collider* or "*v-structure*".

With these two definitions, we can now state the following criteria for Markov equivalence.

Verma-Pearl Criteria for Markov Equivalence:

Two directed acyclic graphs (DAG's) will be Markov equivalent if and only if they have the *same skeleton*, and the *same set of immoralities*.

(Verma and Pearl, 1990, p. 224)

To illustrate the usefulness of the Verma-Pearl criteria for Markov equivalence, let us apply them to a realistic example of a consumer belief structure. For this purpose, consider the expository belief structure for Trader Joe's cereal shown in Figure 12, below. In this case, there is one core product (Trader Joe's cereal) and four brand associations.

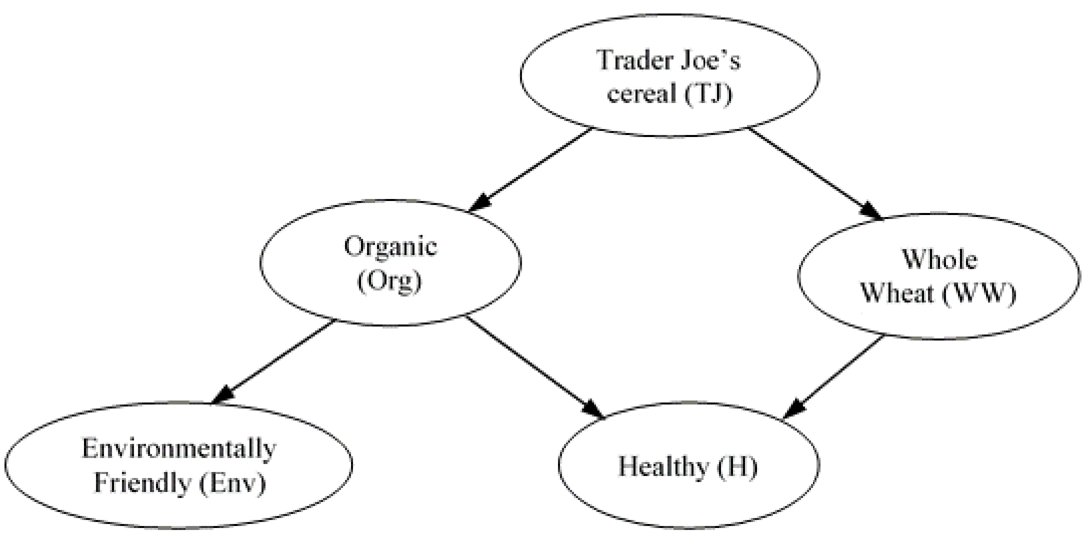


Figure 12. Sample Bayesian network for Trader Joe's cereal

Applying the Verma-Pearl criteria for Markov equivalence to this sample network reveals that there are in fact a set of three additional Bayesian networks that are Markov equivalent to the original network of Figure 12. The full set of Markov equivalent structures is shown in Figure 13, below.[34]

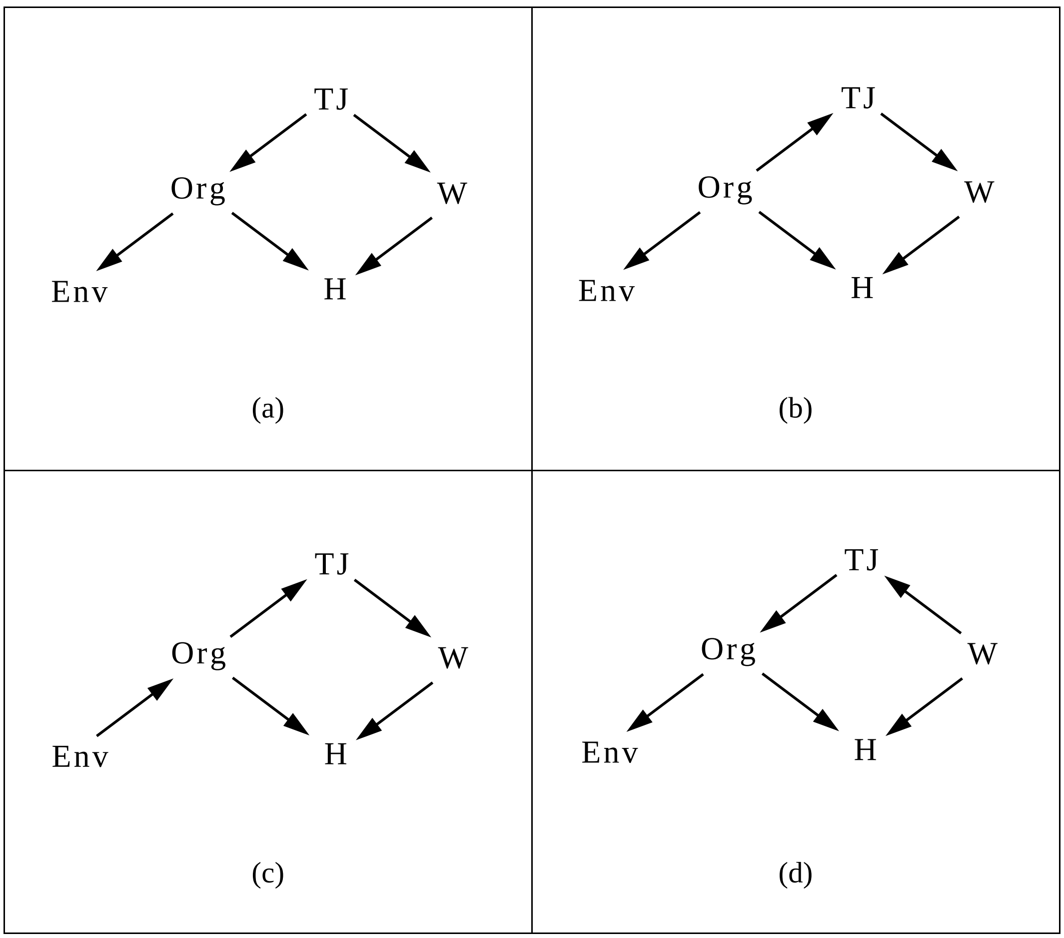


Figure 13. Markov equivalent Bayesian network-based brand concept maps

[34] For reasons of compactness, we are using the abbreviations for each variable within these networks. Please refer to Figure 18 for a complete listing of the full name for each of these variables.

Interestingly, according to the Verma-Pearl criteria, the networks in panels (b), (c), and (d) of Figure 13 are the *only* networks that are Markov equivalent to the original network (which is also shown in panel (a) of Figure 13 for ease of reference). In other words, one cannot simply 'choose' to reverse the direction of a link within the original network or any of these variants. One can check that doing so would either create an immorality (i.e., an uncovered collider structure) that did not exist before, or remove an immorality that already existed in the original network. Either of these outcomes would alter the conditional dependence and independence properties of the network, causing the resulting structure to contain a different set of conditional dependencies and independencies from what is supported by the data that generated the original Bayesian network representation of that domain.

For example, note that the original network (panel (a) of Figure 13) contains precisely one immorality, namely the uncovered collider at variable 'H' ("Healthy"). One can check that each of the other three Markov equivalent structures also contain precisely the same immorality. Further, one can also verify that changing any other link besides the ones that have already been altered to create the structures in panels (b), (c), and (d) of Figure 19 would indeed either destroy the uncovered collider at 'H', or create additional uncovered colliders that were not present in the original structure. As discussed above, such a result would cause the new structure to possess different conditional dependence and independence properties from what would be supported by the data.

As discussed earlier, since all four structures shown in Figure 13 are Markov equivalent, no set of purely observational (i.e., correlational) data will be sufficient to discern any one of these structures from any of the others. If one of these structures is capable of generating a given set of empirical (correlational) data, then the other three structures in this Markov class are equally capable of generating the same set of empirical data.

Interestingly, despite these various structures being observationally equivalent, each can be thought of as representing a different possible causal thought-world for the consumers holding that structure. For instance, the consumers holding structure 13(a) as their belief network for this product can be said to be “Trader Joe’s centric”. In other words, for these consumers, the fact that a cereal is a Trader Joe’s product plays the central role, and the consumer derives all the other properties from this fact. On the other hand, consumers with the belief structure shown in Figure 13(b) are “organic-centric”. For these consumers, the fact that the cereal is an organic product plays the core role, and other facets of the product (such as the fact that it is a Trader Joe’s brand) play a secondary role to the organic nature of the product. Similarly, those consumers holding the belief network of Figure 13(c) are “environmental-centric”, while those in Figure 13(d) are “whole grain centric”.

In terms of potential marketing strategy, note that despite the fact that each of these four structures has the same associative network (i.e., the same undirected link topology, or “skeleton”), each represents a fundamentally different way of viewing the causal nature of the variables within this market. Due to these differences in how these consumers view what causes what within this market, the individuals within each such causal micro-structure are apt to respond completely differently to most marketplace manipulations and promotional efforts as compared to consumers holding any of the other causal micro-structures within this set of observationally equivalent causal structures.

Furthermore, note that if the original network (panel (a) of Figure 13) is the structure which is derived from a reliable set of data, then the Verma-Pearl criteria show that the three additional causal thought-world variants shown in panels (b), (c), and (d) of Figure 13 are the *only* such variants that could exist and still be supported by the collected data. In other words, the principle of Markov equivalence, coupled with the very powerful Verma-Pearl criteria, reveals precisely which cognitive variations can exist based on a given set of data, and further, these criteria rule out any other such variants of these cognitive structures.

It is important to note that were we to consider a purely associative structure (i.e., in which variables become associated, or graphically linked, if there is some appreciable correlation between them, but no directionality is assumed for any of the associations), then all four of these directional variants would collapse into the same exact associative structure, namely the common undirected “skeleton” underlying this general link connection pattern, or topology. In other words, a non-directional technique (such as the BCM technique as it currently exists) would simply coalesce all of these different directional variants into one undirected structure. Therefore, by treating the network as an undirected structure, the potential structural (and hence cognitive) differentiation that is revealed by directional models of this brand concept would be lost, and would be replaced by a potentially less informative structure which would not reveal that several different causal configurations can co-exist within the data (and just as importantly, that any *other* such causal configurations would *not* be supported by the same collected data).

If one further analyzes the collection of Markov equivalent cognitive variants that can co-exist within a directional model of a brand concept network, it becomes apparent that the direction of certain links is 'fixed' (or 'protected') because reversing such a link would either create a new uncovered collider structure that is not supported by the data, or would destroy an existing uncovered collider that is supported by the data. For instance, in all of the cognitive causal variants of the Trader Joe's network in Figure 13, both of the links leading to the brand association "Healthy" are directionally protected since reversing either of them would destroy the uncovered collider at "H" (and in some cases, might even create an additional uncovered collider that is not supported by the data). On the other hand, many of the other links in the structures shown in Figure 13 *are* reversible, in the sense that their reversal merely creates an alternative causal structure that is observationally equivalent to the original structure (i.e., both the original structure and the version that includes the directional reversals would possess the same sets of conditional dependencies and independencies, and hence would be Markov equivalent).

Since the cognitive ramifications of a reversible versus a non-reversible link are so profound, it becomes important to be able to 'summarize' those structural links which are known to be fixed within all variants belonging to a particular Markov equivalence class, and distinguish them from those structural links are known to be variable within that same Markov class. Such a summarization process can be accomplished by creating what is known as an *essential graph* for that Markov equivalence class (e.g., Pearl, 2000).

Within the essential graph of a Markov equivalence class, every link which is directionally stable is shown with that stable directional orientation, and every link which can still be directionally varied while still maintaining the same overall independence properties within that structure (i.e., while still being capable of generating the same set of observed correlations) is shown as an *undirected* link. In this manner, the essential graph for a given Markov equivalence class becomes the standard representative for that class, in that it indicates all of the essential features shared by each causal structure within that class, as well as illustrating the features which are still underdetermined, or variable, within that class.

Given a causal graph G, it is typical to use the notation $[\, G \,]$ to indicate the collection of all causal structures which are Markov equivalent to G and to use the symbol G^* refer to the essential graph for $[\, G \,]$ (Koller and Friedman, 2009). For illustration, the essential graph G^* for the Markov equivalence class of cognitive causal variants $[\, G \,]$ for the Trader Joe's network is shown in Figure 14. (For ease of reference, we have included both the original directed network G and the essential graph G* in this diagram.) Thus we see that a directional technique such as Bayesian network analysis coupled with the Verma-Pearl Markov equivalence criteria is also capable of generating non-directed links (such as those shown in the structure G^* of Figure 14), but does so very carefully, i.e., by first showing that both directions of a particular link can co-exist in the structure without altering the conditional independence properties in the data.

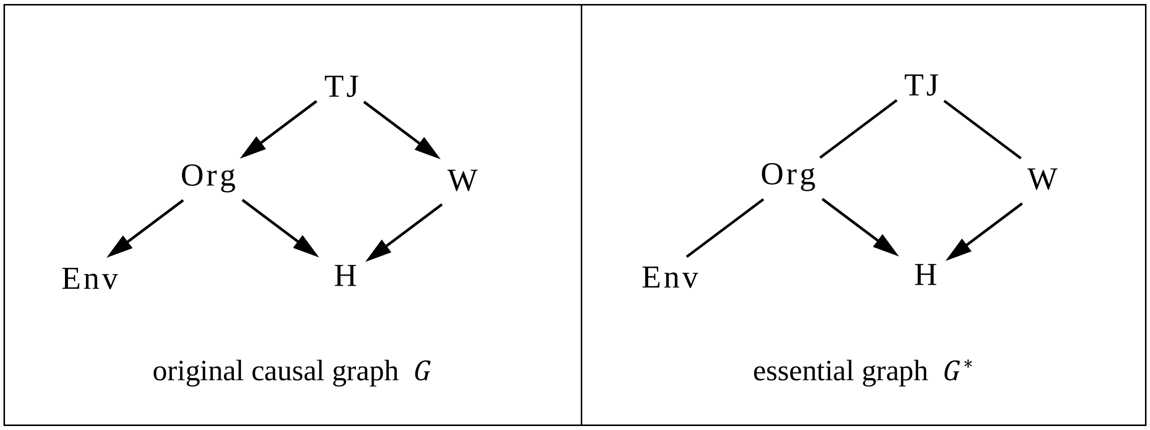


Figure 14. Original causal structure G from Figure 13 (left panel), and the essential graph G^* for the Markov equivalence class $[\,G\,]$ generated by G (right panel).

Based on this essential graph for the equivalence class generated by the original causal structure G, we can see that the observational data which gave rise to G is sufficient to "lock in" two of the directional links in the causal structure which represents that data set. On the other hand, we can see that the remaining three links can be oriented in either direction and still generate a directional structure which can explain the observed data. Based on this representation, for instance, a brand manager for the TJ's brand would have to be careful to note that the only causal directions that are "locked in" by the type of observational data that is available for customers to use in formulating their cognitive structures for these brand associations are the links from *Organic* to *Healthy* and from *Whole Wheat* to *Healthy*. All of the remaining causal directions within the network exist in both directions (within different permissible causal variations of the basic structure), and hence these are the links which the manager will likely find are more easily manipulable (both by the TJ's brand as well as by its competitors).

3. Computational Aspects of Bayesian Network Models

3.1 Bayesian Network Structural Estimation Procedures

Given that the main use of a Bayesian network model of a brand concept network is to understand consumers' reasoning patterns regarding the constituent variables, and that such reasoning patterns may often have an inherent directionality, it stands to reason that the mechanism utilized to uncover the connectivity structure among these constituent variables should be as sensitive to the directional structure as possible. Although multiple forms of structural elicitation exist in the Bayesian networks literature, these methods can roughly be categorized into either *constraint-based* methods, *score-based* methods, or a combination of these two (Koller and Friedman, 2009).

Constraint-based structural elicitation methods seek to list all conditional independencies of the form $A \perp\!\!\!\perp B \mid C$ that exist among disjoint sets A, B, and C of variables within the domain, and then to assemble these independencies into the most likely joint network structure that would exhibit these independencies. Although such approaches are conceptually quite straightforward (and are also quite close in spirit to the intended semantic meaning of the derived network), it is obviously quite difficult in practice to determine all of the conditional independencies in a large data set with enough accuracy to ensure that one has not mistakenly included any spurious independencies, or perhaps neglected to add in certain specific actual independencies that do exist in the domain, but may not have been represented strongly enough in the particular data collected.

As an alternative to such constraint-based structural elicitation methods, certain *structural scores* may be used as a target for optimization, with the idea being that the structure which optimizes such a score given a particular data set will be the most likely network structure for the domain from which that particular data set was collected. Such *score-based techniques* have the advantage of allowing one to essentially "trade-off" imprecision across multiple possible links at the same time, thereby potentially avoiding the "all-or-nothing" nature of the link determination steps that exist within constraint-based methods. In these techniques, the score that one is optimizing is essentially a "stand-in" for the structure that would exist among the variables based on specific conditional independence relationships among those modeled variables.

We will initially examine constraint-based estimation procedures, which are more straightforward and simpler to understand than are score-based techniques. Following this, we will outline the development of structural scores which seek to capture much of the structure of these constraint-based techniques within an abstract function. In addition, as described earlier in this thesis, since Bayesian networks can only be distinguished up to Markov equivalence, these structural estimation procedures are themselves only capable of deriving Bayesian network structure up to a set of Markov equivalent directed structures. The fact that such structural estimation procedures stop at the point at which a Markov equivalence class of structures is determined makes sense from a consumer behavior standpoint, since any member of such a Markov equivalence class represents a cognitive causal variant that is equally capable of representing the observed data, and hence all such cognitive causal variants can co-exist within the respondent population.

3.2 Constraint-Based Structural Estimation

The core concept at the heart of constraint-based Bayesian network structural estimation is that the link structure of a Bayesian network can essentially be thought of as a graphical encoding of the set of conditional independence and dependence relations among the modeled variables which are supported by the given data (Geiger, Verma, and Pearl, 1990; Lauritzen, Dawid, Larsen, and Leimer, 1990). Constraint-based algorithms essentially proceed by first identifying the set of such conditional independence and dependence relationships that exist within the given data (i.e., the set of ‘constraints’ that exist in that data), and then these constraint-based methods attempt to construct a network that best encodes these dependencies and independencies, taking interactions between the derived directions of these relationships into account. Furthermore, since directed (i.e., Bayesian) networks can only be distinguished up to Markov equivalence, these structural estimation procedures are themselves only capable of deriving Bayesian network structure up to a set of Markov equivalent directed structures.

Initially, constraint-based algorithms proceed by enumerating all conditional independence relations of the form $X \perp\!\!\!\perp Y \mid S_{XY}$, where S_{XY} denotes some subset of variables in the domain (not including *X* or *Y*) which serves to render variables *X* and *Y* conditionally independent. Following Kjaerulff and Madsen (2008), the test statistic typically employed for these procedures is the G^2 statistic given by

$$G^2 = 2 \sum_{x,y,z} N_{xyz} \, log\left(\frac{N_{xyz}}{E_{xyz}}\right) \tag{5}$$

where *x* and *y* represent configurations of variables *X* and *Y* respectively, and *z* represents a possible configuration of the 'separator' S_{XY}. Also in this formula, N_{xyz} represents the count of the number of occurrences of the event (*X*=*x* , *Y*=*y* , *Z*=*z*) in the data, with N_{xy}, N_{xz}, and N_z defined similarly. Finally, we define the expected count E_{xyz} via the formula $\left(N_{xy}\ N_{yz}\right) / (N_z)$. Under the null hypothesis of conditional independence of *X* and *Y* given S_{XY}, the overall G^2 statistic will have an asymptotic χ^2 distribution with $(n_X - 1)(n_Y - 1)\prod_{Z\in S_{XY}} n_Z$ degrees of freedom, where n_X , n_Y , and n_Z represent the cardinality of the sample space (i.e., the number of possible configurations) for each of *X*, *Y*, and *Z* respectively. In typical applications, the degree of the separating sets S_{XY} which are utilized in these conditional independence tests must often be limited (usually to three variables or fewer) in order to control the combinatorial expansion of this phase of the procedure, as well as to adjust for the available sample size (since each level of 'nesting' of the conditional independence tests requires successively larger sets of data in order to be reliable).

By repeated application of this initial conditional independence testing procedure, one can determine whether a separating set S_{XY} exists for each pair of variables X and Y in the domain. If no such separator can be found for a particular pair of variables, then these two variables are joined by an undirected link (indicating that they are unconditionally dependent based on the collected data). Once all such unconditional dependence relations have been found, the resulting undirected graphical structure is called the *skeleton* of the network. (In some sense, the skeleton of the network is the structure that the BCM technique derives, assuming that subjects can accurately gauge which variables are directly dependent when eliciting the links in the network.)

Once the network skeleton has been derived, the search for directionality begins via the identification of all collider (i.e., common effect) structures in the network. This stage involves examining each triple of variables {X,Y,Z} for which X and Y are adjacent in the skeleton structure, Y and Z are adjacent in the skeleton structure, but X and Z are not adjacent. For each such triple of variables satisfying this condition, one investigates whether the central variable in the triple (i.e., the variable involved in the two adjacencies, namely the variable Y using the enumeration described above) is a member of any set S_{XZ} which separates the two terminal variables X and Z in that triple. The idea here is that the central variable in a collider structure serves to probabilistically connect the two terminal variables in that structure (viz., Appendix B). Hence, if the central variable Y in the triple {X,Y,Z} is never present in a separator S_{XZ} serving to separate the terminal variables X and Z, it must be the case that the directional structure among these three variables is of the form $X \rightarrow Y \leftarrow Z$. Stated another way, if the directional structure among these variables was any of $X \rightarrow Y \rightarrow Z$, $X \leftarrow Y \rightarrow Z$, or $X \leftarrow Y \leftarrow Z$, then Y would certainly separate X and Z, and hence we would have found variable Y to be a member of at least some separating set S_{XZ} serving to probabilistically separate X and Z.

Once the collider structures within the skeleton have been identified, each additional link within the skeleton that has not yet been directed during the collider identification procedure is examined to determine if its direction can be chosen so as to not introduce any additional colliders or eliminate any colliders that have been previously established.[35] In practice, it may not be possible to orient every structural link in this manner. In some sense, any links which remain unoriented can be thought of as a type of graph-theoretic "confidence region". Furthermore, one can also utilize the Verma-Pearl criteria to investigate whether various structural links are reversible without resulting in removing the derived structure from its Markov equivalence class. Each such additional directed variant within this equivalence class represents a potential causal 'thought-world' or cognitive variation that is consistent with the given data, and which belongs to the equivalence class [G]. Finally, one can determine the essential graph representation G* which represents the entire equivalence class [G] to which the initial solution belongs.

As further described in Spirtes et al. (2000), there are several modifications to this basic constraint-based algorithm, each of which employs certain graph-theoretic properties to simplify or streamline various stages of the procedure. However, all such constraint-based methods follow this same basic heuristic.[36] Overall, the constraint-based methods are very direct, and relatively simple as compared with score-based methods, which are discussed in a succeeding section.

---

[35] Interestingly, Meek (1995) has identified a set of four necessary and sufficient orientation rules which serve to maximally direct the remaining links at this stage of the procedure

[36] For instance, the Peter-Clark (or 'PC') algorithm described in Spirtes et al. (2000) utilizes the fact that if two variables X and Y can be d-separated, then they must be d-separable by either the direct parents of X or the direct parents of Y, which allows one to reduce the search for separating sets to the immediate neighbors of X and Y. In addition, this algorithm improves efficiency by starting with a complete graph on n vertices and removing direct inter-variable links once a separator is found (as opposed to beginning with the empty graph on n vertices and adding direct inter-variable links in cases where no separator can be found).

3.3 Application: Bayesian Network Analysis of the Taco Prima Brand Map

Bayesian network model. In order to illustrate the advantages offered by a Bayesian network analysis of consumers' perceived associations for a well-known brand, data from five hundred individuals regarding their perceptions of the Taco Prima fast food chain were analyzed. Initial interviews revealed that the most common perceived brand associations shared across the respondents were the following six characteristics[37] (each of which is listed with its eventual abbreviation shown at the right):

1. Evokes Good Memories ("Memories")
2. Is a Leader ("Leader")
3. Unique and Different Menu Items ("Unique")
4. Is Innovative ("Innovative")
5. Inexpensive Menu Items ("Inexpensive")
6. Comes Out With New Things ("NewThings")

Since several of these descriptors represent fairly abstract brand characteristics, it might be difficult for respondents to accurately gauge whether their concepts of the Taco Prima brand are characterized by a relation between such descriptors. For example, the brand associations 'Evokes Good Memories' and 'Is a Leader' each do relate to the core brand (Taco Prima), but it may be quite difficult for consumers to be able to judge if their perception of Taco Prima is characterized by an association between these two brand characteristics. On the other hand, recall that Bayesian network structural estimation

[37] Each of these six questionnaire items was collected on a 5-point Likert scale anchored by 1 = 'Strongly Disagree' and 5 = 'Strongly Agree', with the value of 3 representing 'Neither Agree nor Disagree'. The data were obtained from a commercial source with the understanding of privacy regarding the details of the respondent population and the interview procedure.

procedures can operate with simple Likert-style data in which individuals are only asked to express how strongly they associate each attribute with the core brand.[38] In this sense, it is hoped that the Bayesian network-based methodology for determining the detailed interconnection structure among the brand associations within a brand concept map can offer a valuable extension to the pioneering techniques represented by the BCM methodology.

To illustrate the use of a constraint-based structural estimation procedure, we applied the Peter-Clark (or 'PC') algorithm described in Spirtes et al. (2000) to this data, as implemented in the Hugin Lite software package (available at the Hugin Expert website: www.hugin.com). Based on an alpha value of 0.05 for the conditional independence test portion of the PC procedure, the resulting Bayesian network structure for this data is shown in Figure 15, below.

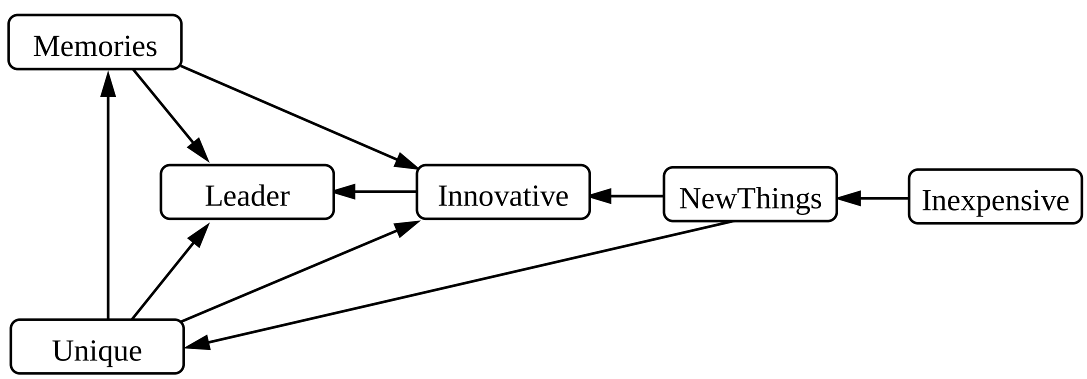


Figure 15. Bayesian network model of the Taco Prima brand concept map

---

[38] The conditional independence properties in the collected data are then used to determine the interconnection structure among the attributes which is most compatible with the dependence and independence properties in that data.

Strategic Analysis. In examining this resulting Bayesian network model, an initial point to notice is that each pairwise connection between variables has a high degree of ‘face validity’. In essence, each individual connection in the structure makes intuitive sense. It is quite interesting that such logical links between variables emerge from a procedure that does not directly probe for those links. Rather, the Bayesian network estimation procedure utilizes respondents’ ratings of how strongly they associate each of the characteristics with the core brand, and then the resulting conditional independence properties in this set of ratings are used to derive the connections between the variables.

Moving on to analyze some of the internal properties of this derived brand concept map, one is initially struck by the fact that the brand association ‘Leader’ is purely a result (or an effect) of other variables, and is not itself predictive of any other variables within this data. This is an interesting finding because it not only shows which specific brand associations lead to consumers perceive Taco Prima as a leader, but it also shows that being perceived of as a leader does not in and of itself ‘help’ (i.e., ‘lead to’) any other of these brand associations. In some sense, it is a ‘terminal’ brand association, in that it leads to no other associations besides itself. This finding, if confirmed, should suggest to Taco Prima that a goal of being perceived of as a leader may not benefit the company as strongly as might have been suspected.

In contrast, the Bayesian network model reveals that being perceived of as ‘Unique’ has the most widespread effect on the other Taco Prima brand associations. Not surprisingly, ‘Unique’ leads to being perceived of as a leader and being thought of as innovative. However, it is quite surprising that 'Unique' leads directly to 'Memories'. In this sense, the perception of Taco Prima as being unique seems to reinforce people's willingness to associate good thoughts or memories with the Taco Prima brand, thereby increasing brand affect as well as elaborative processing of the brand. Hence, according to this Bayesian network derived brand concept map, Taco Prima's most effective point of leverage may in fact be its perception or image of being unique. Furthermore, the only variable which directly influences perceptions of uniqueness is the perception that Taco Prima comes out with new things. Therefore, this fast food chain's most effective promotional strategy may be to continually come out with new products in order to foster the perception of uniqueness, which in turn will encourage strong positive memories of the brand.

Interestingly, a study of Taco Prima's strategic marketing decisions and promotional strategy reveals that this company's management has seemingly come to this same conclusion through trial and error experimentation with the brand. Specifically, the company's product development and marketing strategy seems to center around the constant 'churning' of new inexpensive products, and employs an advertising message that focuses on the novelty, uniqueness, and low prices of these new products. For many large brands such as this, the notion of continually introducing new products has the potential to actually diminish consumers' ability to choose among the firm's offerings, and hence also has the capacity for reducing overall brand evaluations (Broniarczyk et al., 1998; Boatwright & Nunes, 2004). However, as this Bayesian network analysis shows, for Taco Prima the opposite may in fact be true: continually introducing new products seems to increase the perception of uniqueness, which would in turn clearly foster the positive thoughts (i.e., 'memories') which would lead to increased brand equity and increased repeat patronage.

Hence, the directional brand concept map derived through a Bayesian network analysis of consumers' opinions of specific aspects of the Taco Prima brand reveals that the best marketing strategy for this firm is actually quite counterintuitive based on current strategic marketing theories. However, the optimal strategy for this firm which is revealed by the Bayesian network derived brand concept map nonetheless coheres precisely with the successful strategy currently pursued by Taco Prima's management. Note, in addition, that had we utilized current marketing strategy theories to derive a set of presumed relations among these brand associations and then utilized a confirmatory directional procedure such as structural equation modeling in order to test that proposed model, we may never have found such a counterintuitive yet eminently realistic brand concept model such as this.

Cognitive variation. When analyzing the Taco Prima brand image as derived from a Bayesian network model, one must remember that some links may be reversible without altering the independence and dependence properties of the given data, while other links are not reversible in this manner. (Recall that the set of alternative directed structures which are derived from reversing various subsets of those reversible links is called the Markov equivalence class of the original structure.) Based on these notions, we can provide a deeper level of analysis of the brand concept map - one that is only available in directed structures such as this.

In order to facilitate the presentation of this additional layer of analysis, we will use the first letter of each of the Taco Prima brand associations. Since both 'Inexpensive' and 'Innovative' both start with the letter 'I', we will (with due apologies to the Taco Prima Corporation) utilize the mnemonic 'C' (standing for 'Cheap') in place of 'Inexpensive', and we will continue to use 'I' to denote 'Innovative'.

To explore the possible cognitive variants which are observationally equivalent to the original structure, we begin by looking for uncovered collider structures (i.e., 'immoralities') since these are the indicators of the locations of significant conditional independencies in the original data. For ease of reference, the graphical structure of the Taco Prima brand concept map from Figure 15 is reproduced in Figure 16, below, with just the one-letter variable abbreviations used. Examining the network structure in Figure 16 reveals several colliders, but most of them are 'covered' colliders (i.e., the two terminal variables in the collider structure are themselves directly connected). For example, the collider M--L--I is 'covered' by the link from M to I, and the collider N--I--U is 'covered' by the link from N to U, etc. In fact, there is only one uncovered collider in this brand map, namely M--I--N (which is indicated in Figure 16).

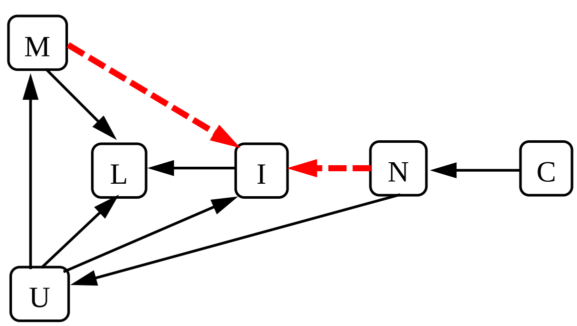


Figure 16. Taco Prima brand concept map, with immorality M--I--N highlighted

Since the collider M--I--N in this network is an immorality, we cannot reverse either the link M → I or the link I ← N since this would ‘destroy’ the immorality, resulting in a structure with different conditional dependence and independence properties from the original. Hence these two links are ‘protected’ in this structure: they can neither be removed nor reversed. Based on this observation, one can see that the link I → L also cannot be reversed because this would create an uncovered collider structure at node I (namely L → I ← N). Furthermore, since the direction of the I → L link cannot be reversed, this means that we cannot reverse the M → L link either, since this would create a cycle among the variables M, L, and I. Hence these two additional links (I → L and M → L) are also directionally protected. Thus we have a ‘core’ of four directionally protected links within this overall structure, as shown in Figure 17 below.[39]

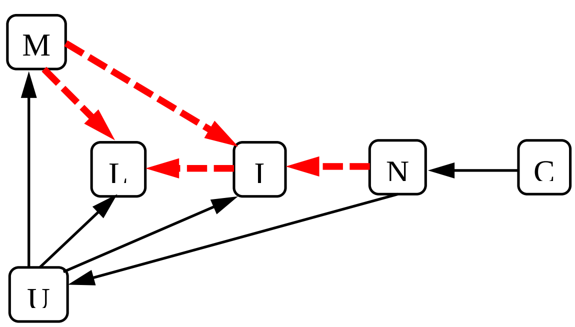


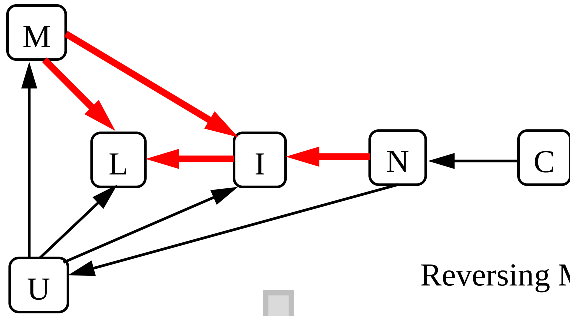
Figure 17. Taco Prima brand concept map with directionally protected links indicated.

[39] In this Bayesian network model, we are disallowing any directed cycles. If one chooses to allow such ‘infinitely repeating cycles’, then the link from M to L would be reversible (and would cause an endless cycle among M, L, and I). However, if all three variables mutually implied one another in an endless cycle or loop, then this would indicate that the combination of all three variables might be a unified aggregate construct rather than three separate constructs. Since consumer behavior considerations suggest that M, L, and I are indeed separate constructs, we will maintain the standard Bayesian network practice of disallowing any configurations which result in a directed cycle.

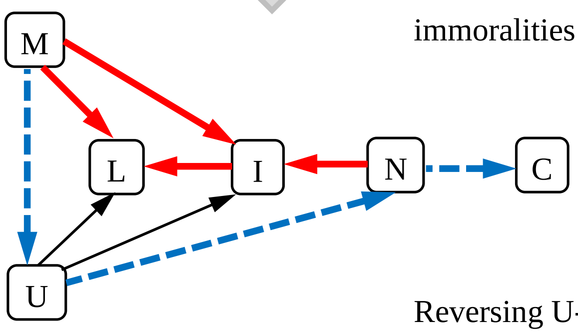
We are now left with only five non-directionally protected links to examine. Of these, the link from C to N is obviously reversible, since it is a terminal link in the structure and its reversal will neither create nor destroy any immoralities, nor will it create any cycles (which are not permitted in a Bayesian network). Hence we now have two rather similar directional variants: the original brand concept map, and the version obtained from reversing the arrow between variables C and N.

Now let us examine the remaining four links case by case. For ease of enumeration, we will work from left to right, beginning with the link between variables U and M. As indicated in Figure 18, reversal of this link would result in an uncovered collider (M -- U -- N) at variable U. Therefore, to accommodate the reversal of U--M, we must also reverse U--N. Of course, this would create a collider structure U -- N -- C, so we also would need to reverse the N--C link. All of these considerations are detailed in Figure 18.

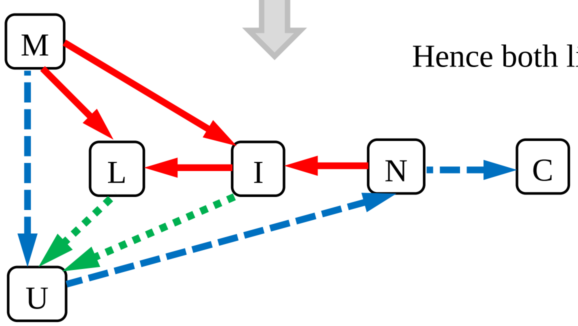
Once the U--M link is reversed (along with the necessary reversals of links U--N and N--C needed to accommodate the initial U--M link reversal), we need to examine the remaining non-protected links for possible additional structural variants that could result from these initial reversals. As indicated in Figure 18, a subsequent reversal of either U--L or U--I is not possible due to the creation of various cycles (see Figure 18 for details). Therefore, the reversal of U--M (along with the required subsequent reversals of U--N and N--C) forms one additional structural variant that is Markov equivalent to the original brand concept map.



Reversing M--U causes U--N and N--C to be reversed, since otherwise we would get immoralities at either U or N.



Reversing U--L without also reversing U--I would lead to a cycle among U, L, and I. Hence both links must be reversed together.



However, reversing both U--L and U--I together causes the new cycle I--U--N to appear, so this arrangement is ruled out.

<u>Conclusion</u>: Once M--U has been reversed (and U--N and N--C are reversed to allow this to occur), no additional changes can be made.

Figure 18. Analysis of the reversal of the M--U link

Moving on to the U--L link, we see that its reversal creates a new immorality at L--U--N and a new cycle at L--U--I. As shown in Figure 19, each of these can be corrected, but these corrections then create a new cycle, and hence are disallowed. Thus, we cannot reverse the link from U to L.

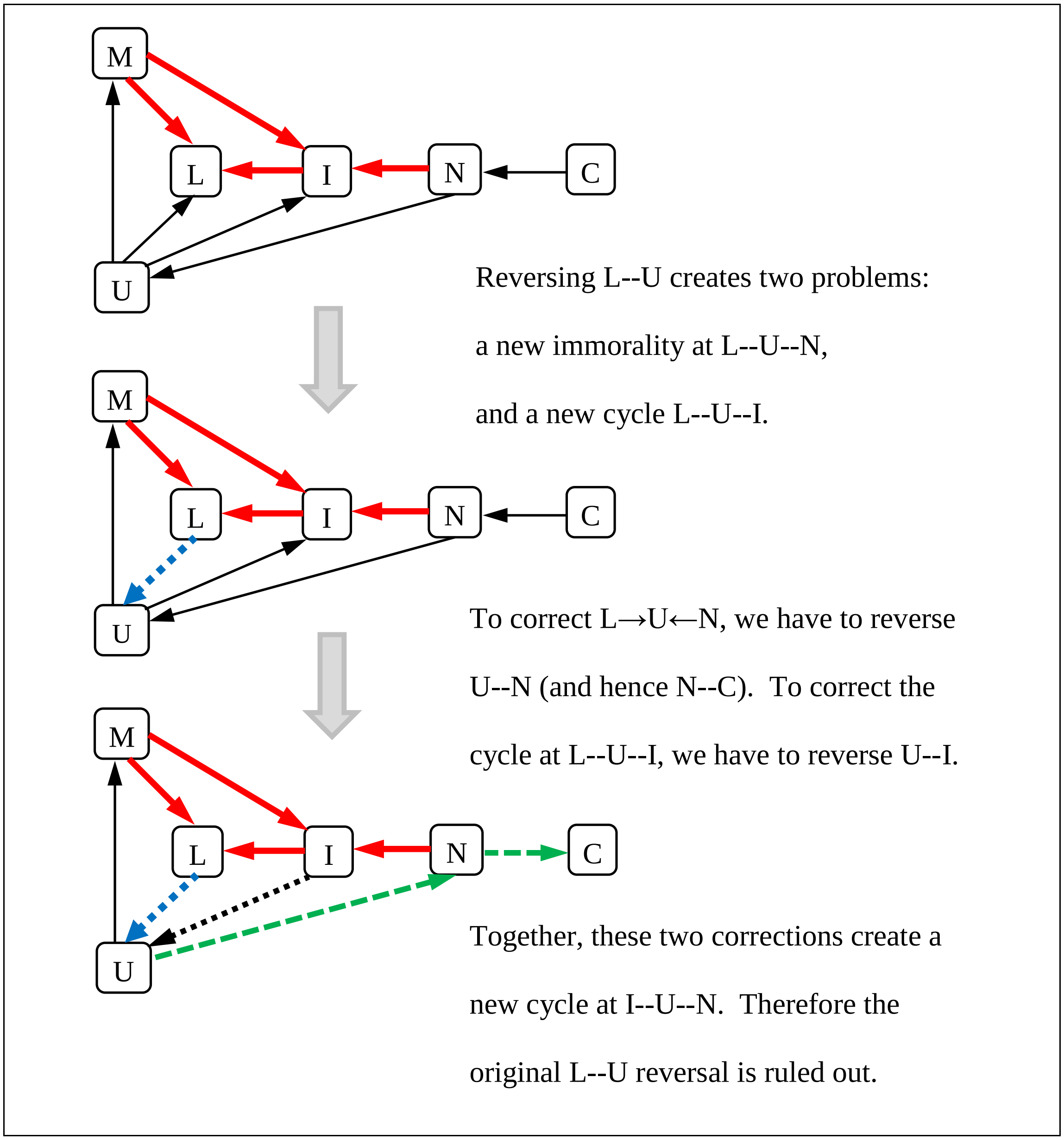
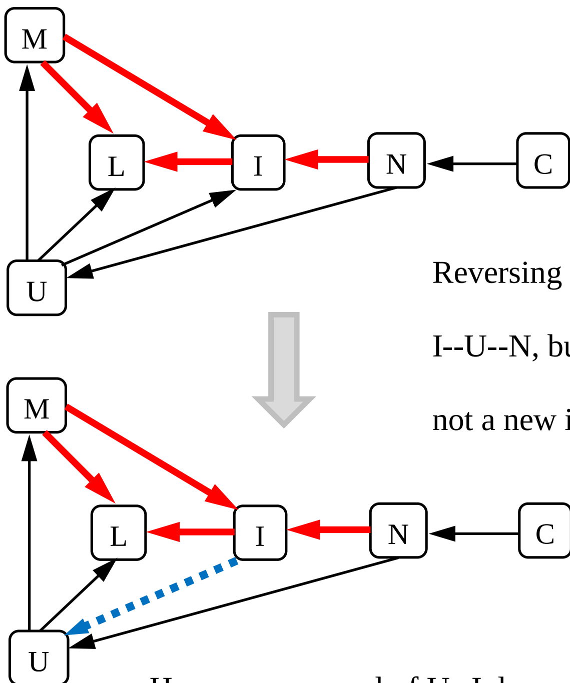


Figure 19. Analysis of the reversal of the U--L link

Now examining the U--I link, Figure 20 shows that its reversal will create a new collider (at I--U--N), but not a new immorality (since that new collider is 'covered' by the N--I link). However, the U--I reversal does create a new cycle, namely the "outside cycle" M--I--U, and removal of that resulting cycle would yield the same structure that we found previously when reversing the M--U link.



Reversing the U--I link creates a new collider I--U--N, but this is a covered collider, so it is not a new immorality.

However, reversal of U--I does create a new cycle at M--I--U. The only way to remove that cycle would be to reverse M--U, but this would create an uncovered collider M--U--N. (We could remove the M--U--N immorality by reversing U--N and N--C, but this would just give us a structure we have already accounted for.) Hence U--I cannot be reversed.

Figure 20. Analysis of the reversal of the U--I link

Finally, as shown in Figure 21, reversal of U--N also entails the reversal of N--C, and then no additional structures are possible once these changes are made.

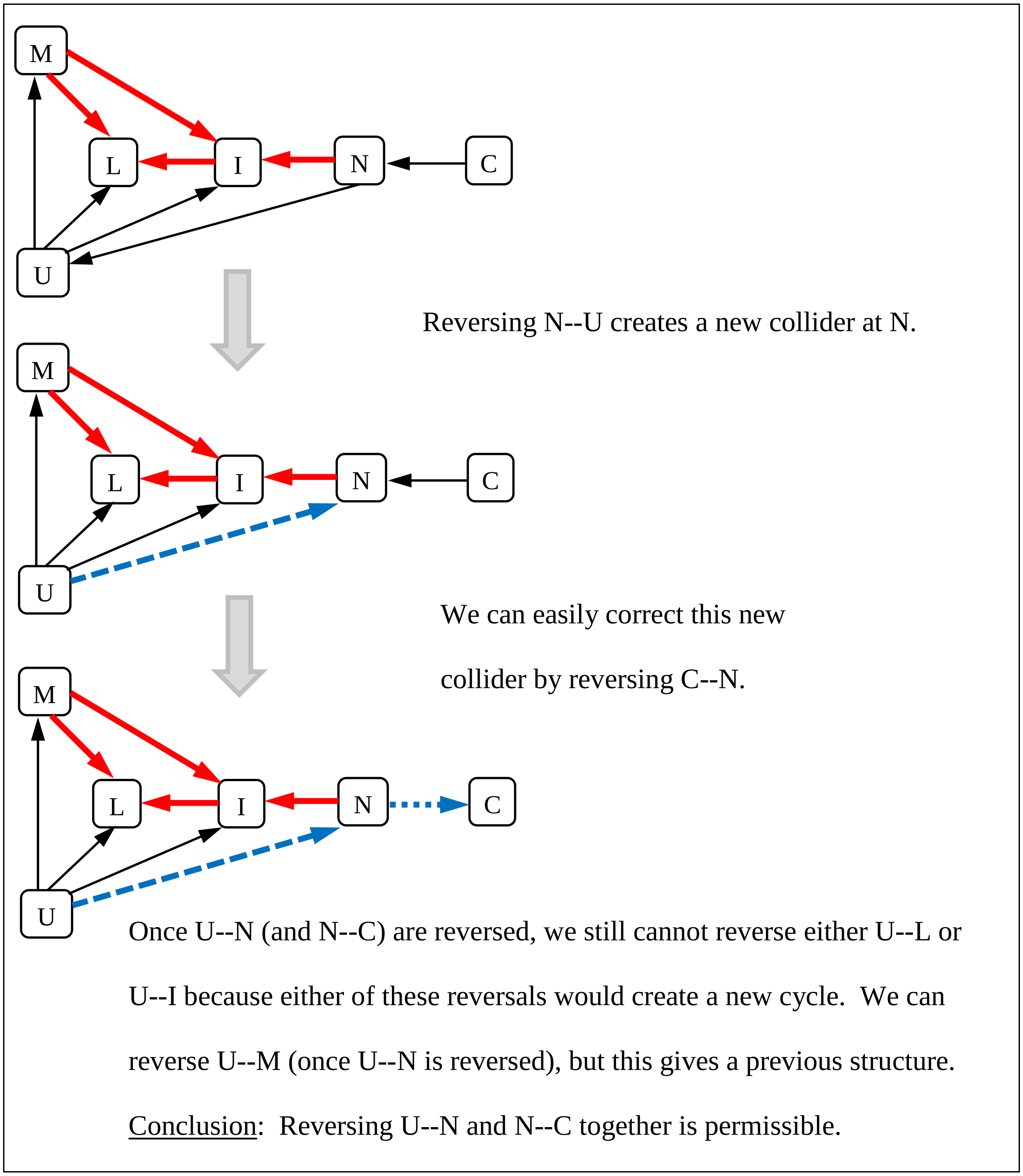


Figure 21. Analysis of the reversal of the U--N link

In summary, a directional analysis of the original Bayesian network structure of Figure 15 reveals that there are three additional cognitive variants of this original structure which are possible given the collected brand concept data. Each of these three additional cognitive variants is Markov equivalent to the original one, and hence all four of these cognitive causal variants can simultaneously co-exist within the data. For reference, these four cognitive variants are shown in Figure 22. (In this Figure, we have continued to show the original directionally protected links in red, while indicating the links which are reversed as dashed blue arrows.) Essentially, these structures all differ from one another along a peripheral pathway which, in the original Bayesian network based brand concept map is oriented as: *Cheap* (i.e., 'Inexpensive') → *New Things* → *Unique* → *Memories*. Within each directional variant (as shown in Figure 22), we see that various subsets of the links on this particular pathway are reversible.

Note that all four of the directional variants in Figure 22 are consistent with the collected data. Hence, according to this data, it is possible for consumers to think that in terms of the Taco Prima brand, 'Cheap' (i.e., 'Inexpensive') implies the existence or creation of new products, which in turn implies uniqueness, and finally generates good memories about the brand. For these consumers, cheap prices are the initial driving force behind this chain of inferences. On the other hand, the data also shows that a chain of reasoning which is 'new product-centered' is also possible, viz., $M \leftarrow U \leftarrow N \rightarrow C$. Similarly, a chain of reasoning that is 'uniqueness-centered' ($M \leftarrow U \rightarrow N \rightarrow C$) is possible, as is a chain of reasoning that is 'memories - centered' ($M \rightarrow U \rightarrow N \rightarrow C$).

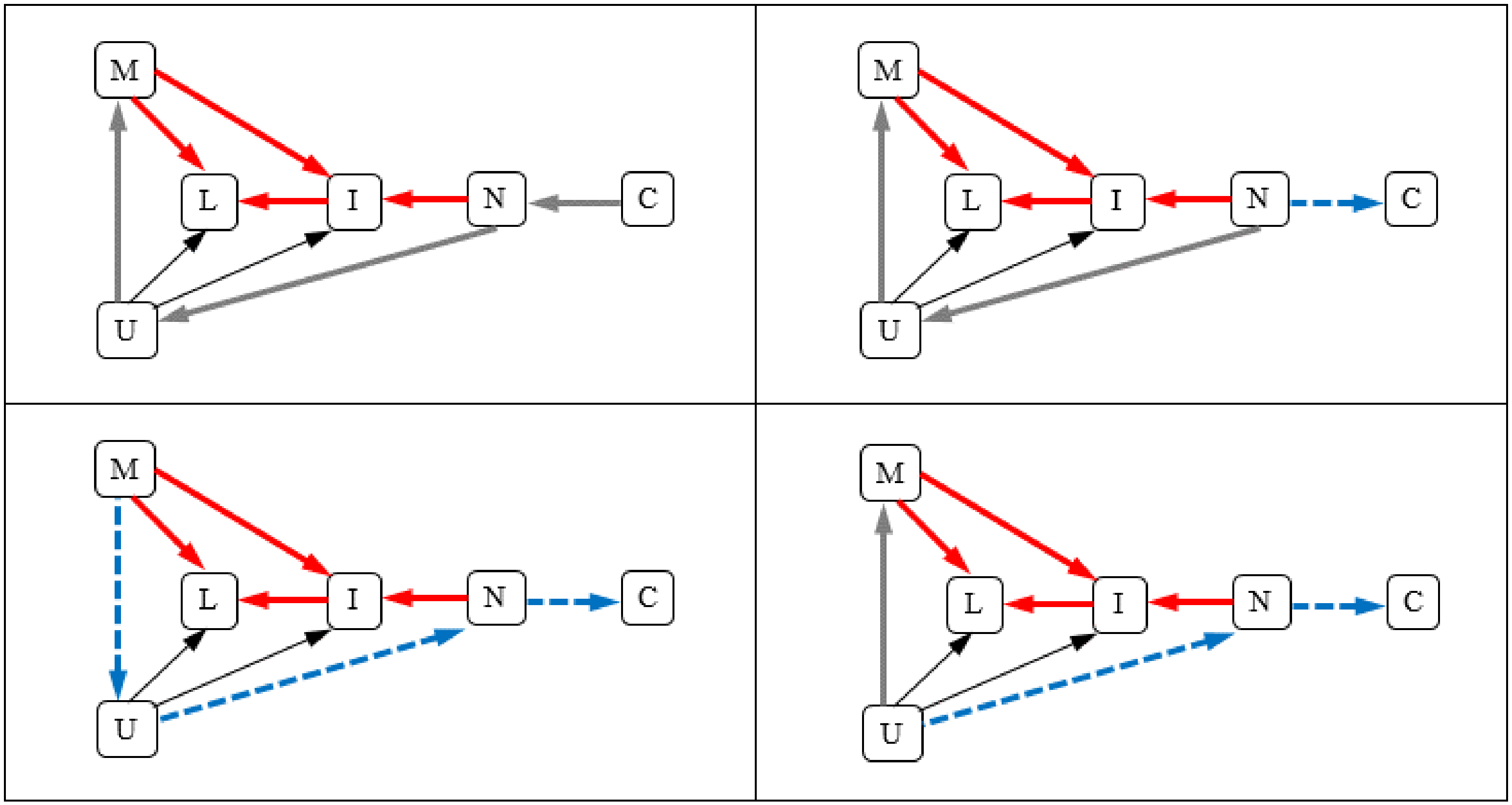


Figure 22. Markov equivalent cognitive variants of the Taco Prima brand concept map

Quite interestingly, all of the other inter-association connections besides the ones on the M ← U ← N ← C pathway are fixed by the data. Hence, as revealed by the Bayesian network analysis of this brand concept map, there are only four cognitive variations which will neither violate any of the conditional independencies in the original data nor generate any additional conditional independencies which are not represented in the data. As an example of these conserved pathways, consider that all structural variants (see Figure 22) must contain the chain of reasoning: *New Things* → *Innovative* → *Leader*. No other connection topologies or connection directions among these three variables would be consistent with the conditional dependence and independence relationships that exist in the data. Similarly, in all structural variants, the brand association 'Leader' is purely a result of other associations (namely 'Memories', 'Unique', and 'Innovative'), and does not in and of itself lead to other brand associations. (One can similarly find multiple other directed relationships that are conserved in all four of the Bayesian network cognitive variants consistent with the original data.)

It is perhaps most intriguing that the portion of this overall brand concept map which is directionally reversible without 'upsetting' the conditional independence properties of the data is precisely the chain of reasoning: *Cheap* (i.e., '*Inexpensive*') → *New Things* → *Unique* → *Memories* which was identified earlier as the most likely driver of product line differentiation and brand equity for this firm. In other words, the Bayesian network based brand concept map for this firm has revealed that *all of the cognitive differentiation among consumers that is consistent with the data must exist along this particular pathway*.

Note that in some sense, this finding can be regarded as a confirmation of the fact that this pathway is the strategically critical one for the firm. If consumers' cognitive structures (i.e., their brand concept maps) did not differ to such a high degree along this pathway, then there might have been enough regularity in the joint distribution of these four variables to allow one or more of them to 'screen off' the others, which we would spot as a non-removable uncovered collider located somewhere along this pathway. Instead, the Taco Prima data supports the existence of four distinct cognitive variations, each of which specifies a slightly different set of predictive and diagnostic relations among the variables on this one particular pathway.

Thus, consumers are apparently highly sensitive to a variety of cognitive influences related to the relationships among the variables *Inexpensive*, *New Things*, *Unique,* and *Memories*. Furthermore, if consumers are so sensitive to these particular relationships for the Taco Prima brand, then they are also likely to be quite sensitive to them when considering any of Taco Prima's direct competitors as well. Hence, it is no surprise that the pathway containing these particular relationships is the strategically critical one for Taco Prima - this is likely to be the "battleground" pathway for firms attempting to compete for a segment of this marketplace niche.

Directed separation analysis. Now consider the 'screening-off' properties that exist among the variables in the Taco Prima brand concept network. By utilizing the formal definition of the d-separation condition (viz., Section 2.3), we can determine the complete set of separation properties that exist within this brand concept structure. Given that this is a relatively small network, the separation properties that it encompasses are relatively limited in number. As it turns out, there are actually six cases within this brand concept network in which probabilistic communication between variables in the network is blocked by the presence of another variable. Of these six directed separation properties, four are relatively obvious from the graphical structure of the network, and two are a bit less obvious. In order to facilitate a detailed discussion of these separation properties, the original Bayesian network-derived brand concept map for Taco Prima is repeated as Figure 23, below.

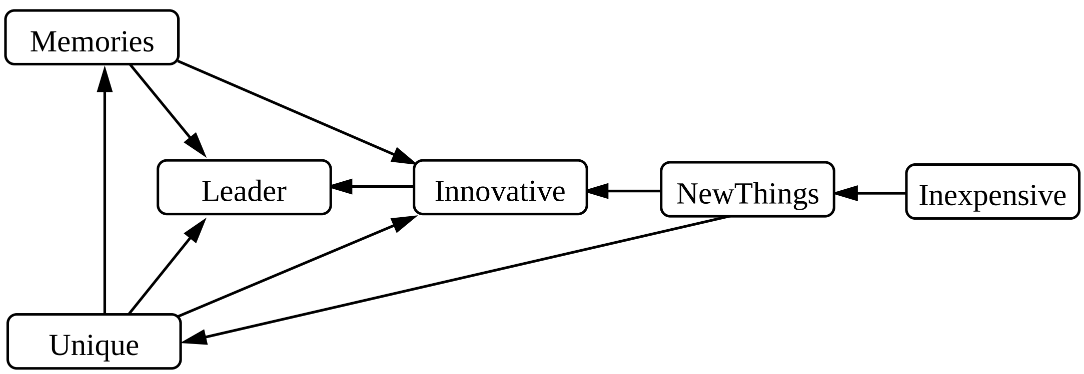


Figure 23. Taco Prima brand concept map (from Figure 15)

To begin with, it should be relatively obvious from the structure of the graph that the brand association *New Things* will act to probabilistically "screen off" each of *Memories*, *Unique*, *Leader,* and *Innovative* from the effects of the brand association *Inexpensive*. These relations can be easily demonstrated because the pathway from *Inexpensive* to any of these four brand associations forms a causal reasoning chain structure involving *New Things* as one of the intermediate variables (and there is no alternate pathway in the network for information to pass from *Inexpensive* to any of these other associations except by passing through New Things).

However, just because a screening-off or directed separation property is obvious from the structure of the network does not mean that it is trivial with regards to the strategic ramifications of the associated brand concept. For instance, these four screening-off relations involving the ability of *New Things* to render the brand characteristic *Inexpensive* irrelevant to the remainder of the network are really quite interesting. Essentially, these four d-separation conditions reveal that the means by which the association *Inexpensive* acts or affects the rest of the Taco Prima brand concept must occur via the action of *Inexpensive* on consumers' perceptions of *New Things*. In other words, the fact that this brand is perceived of as "inexpensive" *must be applied to something* before it can lead to any perceptions of innovativeness, leadership, uniqueness, or positive affect (i.e., 'memories'). Hence, as revealed by the directed separation properties of the Taco Prima brand concept map, the association *Inexpensive* "alone" cannot affect the other brand associations - it can only affect the remainder of the brand concept when it is applied to *New Products*. This finding provides an additional explanation of why such a high rate of "product churning" seems to be such a critical component of Taco Prima's general product development and marketing strategy.

Furthermore, as this brand concept network indicates, the relationship between *Inexpensive* and *New Things* is configured in such a way that once a consumer knows or perceives the value of *New Things* for the Taco Prima brand, that knowledge actually renders *Inexpensive* irrelevant to the remainder of the network. Hence, by and large, consumers seem to already perceive the concept of "New Things" as applied to the Taco Prima brand as encompassing the notion of "Inexpensive". Once a consumer knows or perceives the level of the *New Things* variable for the Taco Prima brand, the notion of "inexpensiveness" provides no additional information that the consumer does not already have.

In addition to these four relatively graphically apparent directed separation properties, the Taco Prima network also possesses two additional d-separation relations that are less obvious to see based merely on the network configuration., namely $(Memories \perp\!\!\!\perp Inexpensive) \mid Unique$ and $(Memories \perp\!\!\!\perp New\ Things) \mid Unique$. One way to derive these facts is to directly employ the causal Markov condition, namely that a variable's parents can render that variable independent of all of its nondescendants in the network. In this case, *Memories* has only one graphical parent, namely *Unique*. Furthermore, *Memories* has two nondescendants in the network, namely *New Things* and *Inexpensive*. Hence by the causal Markov condition, *Unique* (the graphical parent of *Memories*) would have to be capable of rendering *Memories* independent of those nondescendants.

On one level, these two additional separation relationships imply that once consumers know (or assume) the level of uniqueness represented by Taco Prima, the fact that this brand also may offer inexpensive new things no longer affects consumers' affective response to the brand (i.e., 'Memories'). In other words, *New Things* and *Inexpensive* (or, by our earlier discussion, the conglomerate of "inexpensive new things") can only affect the memories (or affective thoughts) that consumers associate with the brand by acting through those consumers' perceptions of the uniqueness of the brand.

However, if we examine the structure of the network a little more closely, we can see that the reason why "inexpensive new things" cannot reach *Memories* through a pathway other than the one passing through *Unique* is because there is a common-effect structure at the brand association *Innovative*. By the properties of common-effect structures (or "colliders"), probabilistic information cannot pass through the node at *Innovative* as long as the level of innovativeness is relatively unknown or vague. On the other hand, if the level of innovativeness was clear to consumers, then information would certainly be able to flow from "inexpensive new things" to *Memories* through the node at *Innovative*, and therefore the brand association *Unique* would no longer be able to d-separate *Innovative* or *New Things* from their effect on *Memories*.

Hence, as it stands, the Taco Prima brand concept network structure raises the very reasonable point that if consumers are unsure of the level of innovativeness represented by the Taco Prima brand, then the fact that this brand may be known for constantly introducing inexpensive new things will affect consumers' "memories" (i.e., affect) for the brand through the role that these inexpensive new things play in establishing perceptions of uniqueness. On the other hand, if consumers *do* perceive a specific level of innovativeness from the Taco Prima brand, then this opens up a second pathway by which consumer feelings or memories for the brand can be affected by the perception of inexpensive new things. Therefore, in general, it is in Taco Prima's best interest to not just foster the perception of constantly coming out with new things, but rather, to also foster the notion that such new things are actually innovative. Thus, it should come as no surprise that this brand's promotional message not only typically contains arguments that the firm's new products are unique, but also typically contains arguments that their constant stream of new products are indeed quite innovative (i.e., the "wow factor" is consistently stressed in their ads in relation to the new products that are introduced). For example, a typical Taco Prima ad might tout the company's new "Doritos taco shell" as not just being unique to Taco Prima, but also as being really clever, e.g., "only Taco Prima could come up with a taco shell made from Doritos", which is a tag line with both uniqueness and cleverness (or innovativeness) connotations.

3.4 Likelihood Scores for Bayesian Networks

The Taco Prima brand concept network analyzed in Section 3.3 was estimated via a constraint-based algorithm. As described earlier, this class of structural estimation procedure directly analyzes each conditional independence relation within the data and constructs a directed network based on distinguishing among the various types of directional three-variable substructures that are found to exist within the recovered structure. The advantage of this method is that it directly analyzes each conditional independence relation in the data and bases its structural estimation on these revered conditional independencies. However, this particular advantage of constraint-based structural estimation procedures can also be regarded as a potential disadvantage with certain data sets, since if there is any lack of certainty about the conditional independencies in the data at the initial stage of structure estimation, then the procedure could possibly result in a non-optimal structure. Of course, this propensity can be controlled to a great degree through analysis of the structures that result from the use of different alpha-values in the conditional independence testing portion of the estimation procedure, but the fact remains that each decision of whether or not two variables in the domain are able to be 'separated' (i.e., rendered conditionally independent of one another) given a third variable (or set of variables) is essentially an 'all-or-nothing' declaration - either the relationship is declared to represent a conditional independence or it is not.

One possible means of correcting for this 'all-or-nothing' determination of whether or not a particular pair of variables is independent conditional on a third variable (or subset of variables) is to utilize some form of comparative scoring function which essentially allows one to 'trade-off' some imprecision between each conditional independence determination. In developing such a measure, we can utilize another aspect of a Bayesian network representation of a set of brand associations, namely the ability of each set of graphical 'parents' in the network to explain (or predict) the distribution of values found in their associated graphical 'child' variable. By seeking the structure which maximizes such a score, one can presumably derive the graphical structure which best exhibits the conditional independence and conditional dependence properties of the data (e.g., Koller and Friedman, 2009). Of course, once again, due to the nature of Bayesian networks, this determination can only be made up to Markov equivalence.

If we let G represent such a graphical structure (i.e., a Bayesian network) under consideration, and we let D represent the specific data set collected from the intended domain of application for the network, a potential choice of 'scoring' function to be optimized is the probability $P(D \mid G)$ of the data set given the potential model under consideration. Since the probability $P(D \mid G)$ represents the *likelihood score* $L(G : D)$, the task of finding an optimal graphical structure G in this manner will also require an optimal choice of parameters θ_G. Therefore the estimation problem can be expressed more completely as one of maximizing $L(\langle G, \theta_G \rangle : D)$. Following Koller and Friedman (2009), we can iterate this optimization in two stages:

$$\max_{G,\theta_G} L(\langle G, \theta_G \rangle : D) \;=\; \max_{G} \left[\max_{\theta_G} L(\langle G, \theta_G \rangle : D) \right] \tag{6}$$

However, since we are assuming that we will choose parameters θ_G which maximize the likelihood of the data given the graph, we can utilize the maximum likelihood parameters $\hat{\theta}_G$, in which case the optimization problem reduces to :

$$\max_{G,\theta_G} L(\langle G, \theta_G \rangle : D) \;=\; \max_{G} \left[L\big(\langle G, \hat{\theta}_G \rangle : D\big) \right] \tag{7}$$

Using $l(\hat{\theta}_G : D)$ to denote the logarithm of this objective likelihood function, we derive the *likelihood score* for the network structure G, viz.,

$$score_L\,(G : D) \;:=\; l(\hat{\theta}_G : D) \tag{8}$$

where once again we are assuming that the MLE parameters $\hat{\theta}_G$ for the graph G are being employed.

To get a better idea of the form of the likelihood score in a realistic scenario, consider a situation in which we would like to compare the scores of two models: $G_0 : \ X , \ Y$ versus $G_1 : \ X \rightarrow Y$. (In other words, we are examining a typical scenario in which we are considering whether or not to add a specific directed link to a basic network structure.) To make matters simple, we will assume that each of X and Y are Bernoulli(θ) random variables. Therefore, once again following Koller and Friedman (2009), the likelihood scores of these two competing models can be computed as:

$$score_L(G_0 : D) = log \prod_m \hat{\theta}_{m(x)} \hat{\theta}_{m(y)}$$

$$= \sum_m \left(log \, \hat{\theta}_{m(x)} + log \, \hat{\theta}_{m(y)} \right) \tag{9}$$

$$score_L(G_1 : D) = log \prod_m \hat{\theta}_{m(x)} \hat{\theta}_{m(y)|m(x)}$$

$$= \sum_m \left(log \, \hat{\theta}_{m(x)} + log \, \hat{\theta}_{m(y)|m(x)} \right) \tag{10}$$

where in each case, $m(x)$ and $m(y)$ represent the m^{th} occurrence (or “realization”) of the variables X and Y respectively within the data stream represented by D.

We can now compare the two likelihood scores as follows :

$$score_L(G_1 : D) - score_L(G_0 : D)$$

$$= \sum_m \left(log\ \hat{\theta}_{m(y)|m(x)} - log\ \hat{\theta}_{m(y)} \right)$$

$$= \sum_{x,y} M(x,y) \cdot log\ \hat{\theta}_{y|x} - \sum_y M(y) \cdot log\ \hat{\theta}_y \qquad (11)$$

where $M(x,y)$ and $M(y)$ represent the number of occurrences of the MLE parameters $\hat{\theta}_{y|x}$ and $\hat{\theta}_y$ respectively in the data set D which consists of M overall observations. Again following Koller and Friedman (2009), we can let $\hat{P}(x,y)$ and $\hat{P}(y)$ represent the respective empirical frequencies within the M-observation data set D, and hence we can rewrite the difference in the likelihood scores of these two models as:

$$score_L(G_1 : D) - score_L(G_0 : D)$$

$$= \sum_{x,y} M\ \hat{P}(x,y) \cdot log\ \hat{P}(y|x) - \sum_y M\ \hat{P}(y) \cdot log\ \hat{P}(y)$$

$$= M \cdot \sum_{x,y} \hat{P}(x,y) \cdot log \frac{\hat{P}(y|x)}{\hat{P}(y)}$$

$$= M \cdot I_{\hat{P}}(X;Y) \qquad (12)$$

where $I_{\hat{P}}(X;Y)$ represents the mutual information between X and Y with reference to the empirical distribution $\hat{P}$ (MacKay, 2003).

The point to be made in this simple case of the choice between model $G_0 : X, Y$ and $G_1 : X \rightarrow Y$ is that the relative increase in the likelihood score that is obtained by adding a link between two previously unconnected variables is directly proportional to the mutual information that exists between those two variables within the empirical distribution. Furthermore, since the mutual information between two variables is a measure of the strength of the dependency between those variables, we see that as the dependency between two variables within a domain increases, the likelihood score measure will increasingly prefer structures that contain an explicit link between those two variables.

We can extend these results to more general Bayesian network configurations as follows:

$$score_L(G : D) = M \cdot \sum_{i=1}^{n} I_{\hat{P}}(X_i ; Pa_{X_i}^G) - M \cdot \sum_{i=1}^{n} H_{\hat{P}}(X_i) \qquad (13)$$

where $Pa_{X_i}^G$ denotes the set of graphical "parents" (direct ancestors) of variable X_i in the structure G, where $I_{\hat{P}}(X_i ; Pa_{X_i}^G)$ represents the mutual information between the i^{th} variable in the model and its graphical parents (computed in the empirical distribution $\hat{P}$), and where $H_{\hat{P}}(X_i)$ is the entropy of the i^{th} variable, i.e., $H_{\hat{P}}(X_i) = \sum_{x_i} \hat{P}(x_i) \cdot log \frac{1}{\hat{P}(x_i)}$. (Of course, the overall enumeration from $i = 1$ to n in equation (13) refers to an enumeration across each of the n variables X_i in the induced model represented by the graphical structure G.)

Generalizing our earlier conclusions from the simple binomial case, we can see from the result in equation (13) that the graphical structure that maximizes the likelihood score (when using the MLE parameters for G) will be the structure for which the parents $Pa_{X_i}^{G}$ of the nodes X_i in the model "explain" the most about their immediate graphical children (in the sense of maximizing the mutual information between those graphical parents and their immediate graphical descendants). In other words, the graphical structure which maximizes the likelihood score with reference to a particular empirical distribution will be the one which possesses a parent-to-child graphical topology which is able to embed the greatest proportion of the overall shared dependence between the variables into the parent-to-child links within that model.

However, this propensity for the likelihood score to prefer models which increasingly place more and more of the overall dependency among the variables into the parent-to-child link structure can often result in serious 'overfitting' of the derived models. As an example, consider the simple case of the comparison between $G_0 : X, Y$ and $G_1 : X \rightarrow Y$. Since the difference in the likelihood scores between these two models is $M \cdot I_{\hat{P}}\ (X; Y)$ and mutual information is always nonnegative (and is strictly positive in all cases in which X and Y are not *exactly* independent within the empirical data), we see that the likelihood score will always prefer the model $G_1 : X \rightarrow Y$ to the model $G_0 : X, Y$ <u>no matter what the data set</u>[40].

[40] This, of course, assumes that we are ruling out the set of cases in which X and Y happen to be *exactly* independent in some empirical data set. However, the set of cases in which this occurs is clearly measure zero within the overall space of possible data sets.

3.5 Bayesian Structural Scores

In order to maintain the desirable features of the likelihood score, but remedy its propensity to severely overfit the data, we can attempt to "spread" the choice of model parameters θ away from a single "concentrated" choice $\hat{\theta}$ and broaden this parameter choice to include more potential values for θ, each "weighted" by some prior likelihood of its occurrence. In other words, we can attempt to remedy the shortcomings of the likelihood score by 'broadening' it to incorporate a *Bayesian* model of the parameter estimation.[41]

To facilitate the development of this broader type of structural measure, consider the overall goal of the overall structural estimation process: namely to maximize the probability $P(G \mid D)$, where G represents a graphical structure (i.e., a Bayesian network) under consideration, and D represents a specific data set collected from the intended domain of application for the network. Now applying Bayes' Theorem, we can rewrite this probability as:

$$P(G \mid D) \;=\; \frac{P(D \mid G) \cdot P(G)}{P(D)} \tag{14}$$

Since the denominator of this expression is essentially a normalizing factor that will be the same for any structure G under consideration, we can restrict our attention to the properties of the numerator. For simplicity, we can linearize the multiplicative form of

[41] This discussion of Bayesian structural scores follows from Koller and Friedman, 2009.

the numerator by taking logs, resulting in an expression that is often called the *Bayesian score* of the considered network G in terms of the given data set D, viz.,

$$score_B\,(G : D) \;:=\; log\; P(D \mid G) \,+\, log\; P(G) \tag{15}$$

The second term in this expression, namely $log\,P(G)$, can be considered as a "structural prior" in the sense that we can assign some prior probability to different initial graphical structures. However, the differences in the probabilities of any reasonable prior structures are typically minimal (Koller and Friedman, 2009), and this term is usually regarded as being negligible compared to the first term in this score function[42]. The first term in Equation 15 is the log of the marginal likelihood of the data (given the structure), which can be computed using standard techniques as follows:

$$P(D \mid G) \;=\; \int_{\theta_G} P(D \mid \langle\, G\,,\, \theta_G \rangle\,)\, P(\theta_G \mid G\,)\, d\theta_G \tag{16}$$

where $P(D \mid \langle\, G\,,\, \theta_G \rangle\,)$ is the likelihood of the data D given the network $\langle\, G\,,\, \theta_G \rangle$, and $P(\theta_G \mid G\,)$ represents our assumptions about the prior distribution of the parameters of the network structure.

Since the marginal log-likelihood of the data, namely $log\;P(D \mid G)$, is the critical quantity in the Bayesian score of a data set, it will be very beneficial to examine its structure in greater detail. To do so, we will make the reasonable assumption that the data are generated by a *multinomial* mechanism, and hence that we can utilize a

[42] For this reason, a *uniform* structural prior is often used in practice.

conjugate Dirichlet prior for the data.[43] To develop a parameterized expression for the marginal log-likelihood of the data, assume that we have a very simple multinomial mechanism: specifically a binomial mechanism with a Beta(α_1, α_0) prior distribution over a single variable X. Letting x^1 and x^0 denote the two possible outcomes of X, and assuming that our data set D consists of M observations, we can write the maximum likelihood function for the data as follows :

$$P(D \mid \hat{\theta}\ ) \quad = \quad \left(\frac{M(x^1)}{M}\right)^{M(x^1)} \cdot \left(\frac{M(x^0)}{M}\right)^{M(x^0)} \tag{17}$$

where $M(x^1)$ and $M(x^0)$ represent the number of occurrences of the outcomes x^1 and x^0 respectively within the M observations that constitute the data set D.

Since the data were generated by a multinomial mechanism, we can compute the likelihood $P(D \mid G)$ of the data given the graph by using a direct computation (rather than via the integral given in equation 16). Specifically, using the chain rule, we can compute the probability of the M observations in the data set D as follows:

$$P(x_1, \dots, x_M) \quad = \quad P(x_1) \cdot P(x_2 \mid x_1) \cdots P(x_M \mid x_1, \dots, x_{M-1}) \tag{18}$$

However, since we are using a Beta(α_1, α_0) prior for X, we can write the successive conditional probabilities in the following general form:

[43] The assumption of multinomial sampling is in keeping with basic Bayesian network practice, and also matches well with typical marketing survey data.

$$P\left(x_{j+1} = x^1 \mid x_1, \ldots, x_j\right) = \frac{j(x^1) + \alpha_1}{j + \alpha} \tag{19a}$$

$$P\left(x_{j+1} = x^0 \mid x_1, \ldots, x_j\right) = \frac{j(x^0) + \alpha_0}{j + \alpha} \tag{19b}$$

where $j(x^1)$ and $j(x^0)$ are the number of occurrences of outcomes x^1 and x^0 respectively within the first j data observations, and where $\alpha = \alpha_1 + \alpha_0$. Therefore, each of the j factors in the probability expression we derive for any particular data string $\left(x_1, \ldots, x_j\right)$ will either be of the form $\frac{\alpha_1 + h}{\alpha + d}$ or $\frac{\alpha_0 + k}{\alpha + d}$ where d represents the observation number being evaluated (i.e., $1 \leq d \leq j$), and where h and k represent the number of x^1 or x^0 outcomes respectively that were seen in the first $d - 1$ data observations within that string.

Since the order of occurrence of the M data observations that constitute the data set D is irrelevant to the overall probability of the particular data string that was observed (i.e., since we are assuming exchangeability of the particular data observations within the string, and hence the overall *numbers* of each particular outcome are the sufficient statistics for that string, rather than the particular sequence of outcome occurrences itself), we can "group" all of the x^1 and x^0 outcomes separately and rewrite the overall probability of the data as :

$$P(x_1, \ldots, x_M) =$$

$$\frac{[\alpha_1 \cdot (\alpha_1 + 1) \cdots (\alpha_1 + M(x^1) - 1)] \cdot [\alpha_0 \cdot (\alpha_0 + 1) \cdots (0 + M(x^0) - 1)]}{\alpha \cdot (\alpha + 1) \cdots (\alpha + M - 1)} \tag{20}$$

Since α_1 and α_0 are hyperparameters for the distribution, they are not necessarily restricted to integer values, and hence we can use the gamma function to rewrite the expression in (20) in a more compact form, viz.,

$$P(x_1, \dots, x_M) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + M)} \cdot \frac{\Gamma\big(\alpha_1 + M(x^1)\big)}{\Gamma(\alpha_1)} \cdot \frac{\Gamma\big(\alpha_0 + M(x^0)\big)}{\Gamma(\alpha_0)} \qquad (21)$$

where we have utilized the relationship $\alpha \cdot (\alpha + 1) \cdots (\alpha + M - 1) = \frac{\Gamma(\alpha+M)}{\Gamma(\alpha)}$ in order to reduce some of the iterated products in the original expression in Equation 20.

Now we can extend this derivation to the case of a multinomial $(x^1, x^2, \dots, x^n)$ distribution over a single variable X with a conjugate Dirichlet$(\alpha_1, \alpha_2, \dots, \alpha_n)$ prior for the data. A parallel derivation to that shown above gives the following analogous expression for the probability of the observed data :

$$P(x_1, \dots, x_M) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + M)} \cdot \prod_{i=1}^{n} \frac{\Gamma\left(\alpha_i + M\big(x^i\big)\right)}{\Gamma(\alpha_i)} \qquad (22)$$

In order to generalize this derivation even further so that it can encompass a full Bayesian network, we must ensure that the same type of modularity (or "locality of influence") that exists in a simple multinomial example over one variable can be replicated in a more complex network structure. We can ensure that such modularity will be maintained in a more general Bayesian network by imposing some regularity conditions, which will take the form of a global independence condition and a local independence condition on the network parameters (Koller and Friedman, 2009).

To begin to derive these regularity conditions, we can first consider how to extend the previous case of a single variable X into the more general case of two variables connected by a directed link $X \rightarrow Y$. In such a case, we will need to jointly model the parameters and the data, such as shown in Figure 24 below (in which the data instances for X and Y are denoted by $X_1, X_2, \dots, X_M$ and $Y_1, Y_2, \dots, Y_M$ respectively).

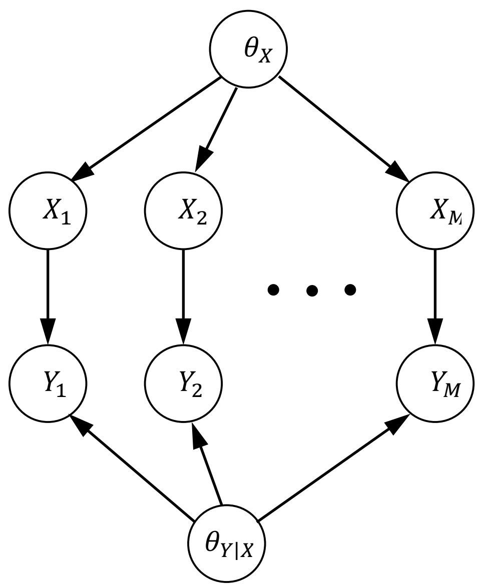


Figure 24. Joint model of I.I.D. data and parameters for a single link $X \rightarrow Y$

From this figure, we can see that knowledge of the parameters θ_X and $\theta_{Y|X}$ will serve to d-separate each pair of linked observations X_i, Y_i from each other such pair of linked observations X_j, Y_j ($j \neq i$). Furthermore, given such a structure, each path between parameters θ_X and $\theta_{Y|X}$ is of the form $\theta_X \rightarrow X_i \rightarrow Y_i \leftarrow \theta_{Y|X}$, and so the parameters θ_X and $\theta_{Y|X}$ will be d-separated from one another given observation of all the $X_i's$ (or, technically, given the *lack* of observation of all the $Y_i's$, which we assume will not occur in a collected data set).

Importantly, under such a data structure, *if the parameters θ_X and $\theta_{Y|X}$ were independent a priori, they will remain independent once updated by the network*. In other words, a network structure of this form will not induce any network-derived dependencies between a priori independent parameters. Furthermore, under such a structure, it is possible to compute the posterior distributions over θ_X and $\theta_{Y|X}$ independently of each other. This independence property of the parameters is what we want to "mirror" within more complex Bayesian networks. More specifically, for a general Bayesian network with parameters $\theta = \left(\theta_{X_1|Pa_{X_1}}, \dots, \theta_{X_n|Pa_{X_n}}\right)$ we will say that the prior $P(\theta)$ possesses *global parameter independence* if it can be decomposed into a product $\prod_i P\left(\theta_{X_i|Pa_{X_i}}\right)$. Most standard priors will satisfy such a criterion, but if, for instance, it is assumed that certain nodes in the network "share" certain properties, such as the same propensity to assign certain values to observations, then such a global parameter independence criterion would fail to hold.

In the setting of a general Bayesian network, the benefit of having a prior that satisfies global parameter independence can be further seen by analyzing the expression for the posterior distribution of the parameter:

$$P(\theta \mid D) = \frac{P(D \mid \theta) \cdot P(\theta)}{P(D)} \tag{23}$$

In this expression, the $P(D \mid \theta)$ term is the (global) likelihood function for the network, and general results from Bayesian network theory (e.g., Koller and Friedman, 2009) show that it can be decomposed into a product of local likelihoods:

$$
\begin{aligned}
P(\, D \mid \theta \,) \quad &= \quad \prod_i L_i \left(\theta_{X_i \mid Pa_{X_i}} : D \right) \\
&= \quad \prod_i \prod_m P \left(\, (x_i)_m \,\middle|\, \left(Pa_{X_i} \right)_m \; : \; \theta_{X_i \mid Pa_{X_i}} \right)
\end{aligned}
\qquad (24)
$$

Since the marginal likelihood term $P(D)$ in the denominator of the posterior distribution $P(\theta \mid D)$ is a normalizing factor (and additionally does not depend on G), the only remaining term to consider from the posterior distribution expression (23) is the expression for the prior distribution of the parameters. This is where the assumption of global parameter independence becomes useful, since under this assumption, we can express $P(\theta)$ as $\prod_i P\left(\theta_{X_i \mid Pa_{X_i}}\right)$ and hence the overall posterior for θ becomes:

$$
\begin{aligned}
P(\, \theta \mid D \,) \quad &= \quad \frac{P(\, D \mid \theta \,) \cdot P(\theta)}{P(D)} \\
&= \quad \frac{\prod_i \; L_i \left(\theta_{X_i \mid Pa_{X_i}} : D \right) \; \prod_i P \left(\theta_{X_i \mid Pa_{X_i}} \right)}{P(D)} \\
&= \quad \frac{1}{P(D)} \prod_i \left[L_i \left(\theta_{X_i \mid Pa_{X_i}} : D \right) \cdot P \left(\theta_{X_i \mid Pa_{X_i}} \right) \right]
\end{aligned}
\qquad (25)
$$

As Equation 25 shows, each component $\theta_{X_i|Pa_{X_i}}$ of θ only contributes to one factor within the overall product expression for the posterior, and hence under this assumption of global parameter independence of the prior $P(\theta)$, we can decompose the *global* posterior distribution into a product of *local* posterior distributions, viz., $P(\ \theta\mid D\ )\ = \prod_i\ P\left(\ \theta_{X_i|Pa_{X_i}}\ \middle|\ D\ \right)$.

As shown above, the global parameter independence condition allows us to reduce the global Bayesian estimation problem to a series of far simpler local Bayesian estimation problems. However, within these local estimation problems, we will typically wish to assume a particular form of *local* parameter independence. To derive this condition, once again consider the basic case of a single directed link $X \rightarrow Y$ between two network variables. In order to analyze the specific local properties of this structure, we will assume a simple binary distribution on each of these two variables, and a tabular CPD for representing their joint distribution. In essence, this assumption amounts to adding a local model to the global structure shown in Figure 22.

In this model for the structure $X \rightarrow Y$, the binary variable X is relatively easy to analyze: if we assume that we have a Dirichlet prior for its parameter θ_X, then the posterior distribution of θ_X given the data $x_1, \ldots, x_M$, namely $P(\theta_X \mid x_1, \ldots, x_M)$, will once again be Dirichlet. However, the remaining parameters are the conditional ones, namely the vector $\theta_{Y|X} = \theta_{y^0|x^0}, \ldots, \theta_{y^1|x^1}$. If we assume that the influence of each of x^0 and x^1 are independent a priori, then we can use a separate Dirichlet prior for each of $\theta_{Y|x^0}$ and $\theta_{Y|x^1}$, yielding the overall prior $P\left(\theta_{Y|X}\right) = P\left(\theta_{Y|x^0}\right) \cdot P\left(\theta_{Y|x^1}\right)$. Given these assumptions, the modified version of Figure 22, in which we have added the local parameterizations to the model, would appear as shown in Figure 25, below (where we are once again using $X_1, X_2, \ldots, X_M$ and $Y_1, Y_2, \ldots, Y_M$ to denote the various data instances).

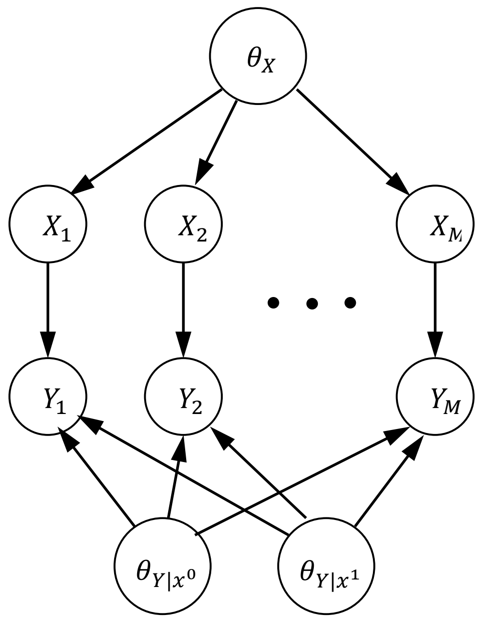


Figure 25. Joint model of I.I.D. binary data and parameters for a link $X \rightarrow Y$

Interestingly, it would appear that observation of the data on Y would serve to correlate the two conditional parameters $\theta_{Y|x^0}$ and $\theta_{Y|x^1}$ (since observation of the central node Y_i in a collider such as $\theta_{Y|x^0} \rightarrow Y_i \leftarrow \theta_{Y|x^1}$ will allow the two terminal variables to probabilistically influence each other). However, the "choice" of which conditional parameter $\theta_{Y|x^0}$ or $\theta_{Y|x^1}$ actively governs the probability of a specific data instance on Y is actually a function of which particular outcome (x^0 or x^1) was observed for X during that particular realization of the network data.

Specifically, if the i^{th} observation X_i of variable X happens to show value x^0 , then the conditional distribution of observation Y_i for variable Y will be governed by parameter $\theta_{Y|x^0}$, and the link from $\theta_{Y|x^1}$ to Y_i in the collider $\theta_{Y|x^0} \rightarrow Y_i \leftarrow \theta_{Y|x^1}$ will no longer be active. In symmetric fashion, of course, the link from $\theta_{Y|x^1}$ to Y_i in the collider structure $\theta_{Y|x^0} \rightarrow Y_i \leftarrow \theta_{Y|x^1}$ will become inactive whenever the i^{th} observation X_i of variable X happens to show value x^1. Therefore, for each realization of the data for these variables, we have a form of context-specific independence of $\theta_{Y|x^0}$ and $\theta_{Y|x^1}$ given the particular value of the variable X that was observed, namely the relation $P\left(\theta_{Y|X} \mid D \right) = P\left(\theta_{Y|x^0} \mid D \right) \cdot P\left(\theta_{Y|x^1} \mid D \right)$. Based on this analysis, we can show that if the priors $P\left(\theta_{Y|x^0}\right)$ and $P\left(\theta_{Y|x^1}\right)$ are Dirichlet with hyperparameters $\left(\alpha_{y^0|x^0} ,\ \alpha_{y^1|x^0}\right)$ and $\left(\alpha_{y^0|x^1} ,\ \alpha_{y^1|x^1}\right)$ respectively, then the posterior distributions $P\left(\theta_{Y|x^0} \mid D \right)$ and $P\left(\theta_{Y|x^1} \mid D \right)$ will be also be Dirichlet with updated hyperparameters $\left(\alpha_{y^0|x^0} + M[x^0, y^0] \ , \alpha_{y^1|x^0} + M[x^0, y^1] \right)$ and $\left(\alpha_{y^0|x^1} + M[x^1, y^0] \ , \alpha_{y^1|x^1} + M[x^1, y^1] \right)$ respectively (Koller and Friedman, 2009).

We now wish to apply this form of network decomposition to a more general Bayesian network structure. As was the case above, we wish to be able to guarantee that if we have Dirichlet priors for the parameters of our network, that we will be able to once again recover Dirichlet posterior parameter distributions from that network. First, we need to extend our previous analysis of a network variable Y and its graphical parent X, and instead consider the more general situation of a variable X and its set of graphical parents U. As we did above, we will wish to assume that the prior distribution $P(\theta_{X|U})$

over the conditional parameter $\theta_{X|U}$ can be written as a product of local priors involving the graphical parents of X. Therefore, we will extend our two-variable local independence assumption $P(\theta_{Y|X}) = P(\theta_{Y|x^0}) \cdot P(\theta_{Y|x^1})$ by instead writing this assumption as $P(\theta_{X|U}) = \prod_u P(\theta_{X|u})$.

Once again, as was done in the simpler case of a single directed link $X \rightarrow Y$, it will typically be beneficial to "couple" the two assumptions of global and local parameter independence. As described above, global parameter independence asserts that the global prior $P(\theta)$ can be decomposed into a product $\prod_i P\left(\theta_{X_i|Pa_{X_i}}\right)$ of local priors for the parameters corresponding to each variable X_i conditional on its set of graphical parents in the network. Following this, the assumption of local parameter independence will allow us to write each local prior $P(\theta_{X|U})$ as a product $\prod_u P(\theta_{X|u})$ of priors for X conditional on each individual parent $u \in U$.

In analogy to the procedure used in the previous example, once we assume both the global and the local parameter independence conditions for the prior $P(\theta)$, we can extend the previous analysis to show that the posterior distribution for the parameters will also have a local decomposition, given by Equation 26 (e.g., Koller and Friedman, 2009).

$$P(\,\theta \mid D\,) \quad = \quad \prod_i \prod_{Pa_{X_i}} P\left(\,\theta_{X_i|Pa_{X_i}} \,\middle|\, D\,\right) \tag{26}$$

Based on this local decomposition, we can also show that if each $P(\theta_{X|u})$ is Dirichlet with hyperparameters $(\alpha_{x^1|u}, \ldots, \alpha_{x^n|u})$, then the posterior distribution $P(\theta_{X|u} \mid D)$ will be Dirichlet with hyperparameters $(\alpha_{x^1|u} + M[u, x^1], \ldots, \alpha_{x^n|u} + M[u, x^n])$.

Now we can apply these results to the estimation of the Bayesian score for a given network. Recall that this score is derived from the numerator of the expression for the posterior probability of a graph given the observed data, namely $P(G \mid D) = \frac{P(D \mid G) \cdot P(G)}{P(D)}$ Specifically, $score_B\,(G : D)$ is defined as the log of this numerator, i.e., $score_B\,(G : D) \;:=\; \log\, P(D \mid G) \,+\, \log\, P(G)$. Since the structural prior $P(G)$ typically makes a negligible contribution to the score, we instead focused upon the term containing the marginal likelihood of the data given the graph structure, namely $P(D \mid G)$, which can be computed as $P(D \mid G) = \int_{\theta_G} P(D \mid \langle\, G\,,\, \theta_G \rangle\,)\, P(\theta_G \mid G\,)\, d\theta_G$.

Our goal is to further characterize this integral by applying the global and the local parameter independence assumptions for the prior. Under the global parameter independence assumption only, this integral can be expressed as:

$$P(\, D \mid G\,) =$$

$$\prod_i \int_{\theta_{X_i|Pa_{X_i}}} \prod_m P\left(\, (x_i)_m \,\middle|\, \left(Pa_{X_i}\right)_m,\ \theta_{X_i|Pa_{X_i}}, G\,\right) P\left(\, \theta_{X_i|Pa_{X_i}} \middle|\, G\right) d\theta_{X_i|Pa_{X_i}} \qquad (27)$$

In addition, if we also maintain the local parameter independence assumption, we can further break this expression down “parent-by-parent” within the parental set U, giving us the expanded integrand sown in Equation 28:

$P(\, D \mid G \,) \ =$

$$\prod_{i} \prod_{u_i \,\in\, Val\left(Pa_{X_i}\right)} \int_{\theta_{X_i|u_i}} \prod_{m \,:\, (u_i)_m = u_i} P\big(\, (X_i)_m \,\big|\, u_i \,,\, \theta_{X_i|u_i}, G \,\big)\, P\big(\, \theta_{X_i|u_i} \big|\, G \big) d\theta_{X_i|u_i} \qquad (28)$$

Now recall that in the case of a multinomial $(x^1, x^2, \ldots, x^n\,)$ distribution over a single variable X with a conjugate Dirichlet$(\alpha_1\,, \alpha_2\,, \ldots\,,\ \alpha_n)$ prior for the data, we were able to show that the probability of the data reduced to the expression $P(x_1\,, \ldots, x_M) \ = \frac{\Gamma(\alpha)}{\Gamma(\alpha+M)} \cdot \prod_{i=1}^{n} \frac{\Gamma\left(\alpha_i + M(x^i)\right)}{\Gamma(\alpha_i)}$. In the more general case of a Bayesian network with graphical structure G, we will similarly be able to use the closed-form expression for the marginal likelihood of the Dirichlet prior in order to derive an analogous expression for the probability of the data given the graphical structure of the network (Koller and Friedman, 2009):

$P(\, D \mid G \,) \ =$

$$\prod_{i} \prod_{u_i \,\in\, Pa_{X_i}} \frac{\Gamma\left(\alpha_{X_i|u_i}\right)}{\Gamma\left(\alpha_{X_i|u_i} + \ M[u_i]\,\right)} \cdot \prod_{x_i^j \,\in\, Val(X_i)}^{n} \frac{\Gamma\left(\alpha_{x_i^j|u_i} + M\left[x_i^j, u_i\right]\right)}{\Gamma\left(\alpha_{x_i^j|u_i}\right)} \qquad (29)$$

where $\left\{ \alpha_{x_i^j|u_i} :\ \ j = 1\,, \ldots, |X_i| \right\}$ are the hyperparameters for $P\left(\theta_{X_i|Pa_{X_i}} \middle| \, G \right)$, and where $\alpha_{X_i|u_i} = \ \sum_j \ \alpha_{x_i^j|u_i}$ are the summed “pseudo-counts” associated with the respective Dirichlet hyperparameters.

Overall, the Bayesian score for a Bayesian network with Dirichlet priors will be of the form: $score_B(G : D) := \log P(D \mid G) + \log P(G)$, where the $P(D \mid G)$ term will have the closed-form expression given in Equation 29, and where the structural prior $P(G)$ will typically be either uniform or close to uniform (and will typically play a negligible role compared to the $P(D \mid G)$ term).

Asymptotically as the size M of the data set goes to infinity, the Bayesian score does approach the relatively simpler Bayesian Information Criterion (BIC) measure given by $score_{BIC}(G : D) := l(\hat{\theta}_G : D) - \frac{\log M}{2} \cdot Dim[G]$, where the $\hat{\theta}_G$ are the MLE parameters for the network, and $Dim[G]$ measures the number of independent parameters in the network (MacKay, 2003; Koller and Friedman, 2009). Furthermore, it can be shown that both the BIC and Bayesian measures are *consistent* up to Markov equivalence (i.e., both measures will converge to some member of the Markov equivalence class of the "true" generating structure G^* as the size M of the data set goes to infinity).

Therefore, in cases where one merely wishes to utilize the derived graphical structure to compute marginals over various queries, it might seemingly suffice to derive the graphical structure by optimizing any particular consistent scoring function. However, when one needs to allow for *directionality* within the derived structure to play a role, this entails the use of the structure to potentially compute the results of interventions to the graph, as well as to distinguish between which direction represents predictive reasoning in the mind of the consumers, versus which direction represents diagnostic reasoning for those consumers. Since such queries are particularly sensitive to the specific "microstructure" of the network, one needs to be sure that the structural measure that most directly reflects that microstructure is utilized in the structural estimation procedure. Hence, for Bayesian network modeling tasks in which such microstructural and directional considerations are likely to prove critical, it may be safest to utilize either the Bayesian measure, or the closely related Bayesian Information Criterion (BIC) measure.

What is especially interesting about the Bayesian structural score itself is that it is able to be a consistent estimator of the generating distribution *without requiring any explicit 'penalizing' function for structural complexity*. Compare this with other consistent structural estimators derived from information-based (i.e., entropy or minimum description length) concerns, such as the Bayesian Information Criterion (BIC) score, for example. Recall that the Bayesian Information Criterion measure is given by the function: $score_{BIC}(G : D) := l(\hat{\theta}_G : D) - \frac{\log M}{2} \cdot Dim[G]$, where $\hat{\theta}_G$ are the MLE parameters for the network, M represents the cardinality of the data on which the estimate is made, and $Dim[G]$ measures the number of independent parameters in the derived network. This score can be re-written to more explicitly represent its information-theoretic foundations as follows (Koller and Friedman, 2009):

$$score_{BIC}(G:D) := M\sum_{i=1}^{n}\boldsymbol{I}_{\hat{P}}\left(X_i ;\, Pa_{X_i}\right) - M\sum_{i=1}^{n}H_{\hat{P}}(X_i) - \frac{\log M}{2}Dim[G] \tag{30}$$

The entropy term in this expression (i.e., the second term on the right-hand side) is not sensitive to the specific graphical structure of the underlying network, and hence this term will not influence the choice of network structure. This leaves a trade-off between the positive contribution of the mutual information term $M \sum_{i=1}^{n} \boldsymbol{I}_{\hat{P}}\left(X_i ;\ Pa_{X_i} \right)$ and the negative contribution of the model complexity term $\frac{log\, M}{2} Dim[G]$. Therefore, a measure such as $M \sum_{i=1}^{n} \boldsymbol{I}_{\hat{P}}\left(X_i ;\ Pa_{X_i} \right)$ which is based purely on the shared information between the various nodes of the network and their graphical 'parents' will need to be augmented by a complexity-penalizing term in order to avoid overfitting. On the other hand, the Bayesian measure (namely, the measure given by $score_B\,(G : D)\ :=$ $\log\ P(D \mid G)\ +\ \log\ P(G)$, where the $P(D \mid G)$ term is given by the closed-form expression given in Equation 29, and where the structural prior $P(G)$ will typically be either uniform or close to uniform) is able to be a consistent structural estimator without the need for introducing such a complexity-penalizing term because the Bayesian measure simply incorporates the precise independence relationships that would be expected from multinomial data within a directed graphical structure which obeys the specific local and global parameter independence properties that are quite natural to expect in such circumstances. In other words, although other measures such as the B.I.C. criterion can also aggregate individual data into a joint structure, it does not accomplish this task in as natural of a manner as the Bayesian structural measure.

Furthermore, since the mutual information term within the B.I.C. measure is a linear function of the data size M while the model complexity "penalty" term is a logarithmic function of M, the complexity penalty term will have its maximum relative contribution (relative to the contribution of the mutual information, or "model-fit" term) when the size of the data set is small (Koller and Friedman, 2009). Hence, with moderately-sized data sets, the structural differences between the independence-derived Bayesian score and the information-theoretic B.I.C. score may prove to be important, and one may wish to utilize the Bayesian score in such cases. This is especially true in cases where one wishes to maximally discern differences between the predictive and diagnostic directions within the brand concept map of the targeted population.[44]

Additionally, one should note that the B.I.C. measure converges to the Bayesian measure in the large-sample limit. Since in the large-sample limit, the mutual information term $M \sum_{i=1}^{n} \boldsymbol{I}_{\hat{P}}\left(X_i \, ; \, Pa_{X_i} \right)$ of the B.I.C. measure vastly predominates, this implies that the Bayesian measure must also be maximizing the ability of the parents of each node to explain the values of their child nodes (Koller and Friedman, 2009). Within the B.I.C. measure, the relative contributions of the complexity penalty and the mutual information term vary as the sample size grows. However, the structure of the Bayesian measure remains constant over increasing sample sizes. Hence the Bayesian measure's propensity to maximize the parent-to-child predictiveness must be occurring at all different sample sizes. Therefore for multinomial data (which is the predominant form of data encountered in typical survey-based marketing research), the Bayesian measure represents an essentially optimal means of aggregating the subjects' individual data in such a way as to consistently derive structures in which the parents of each node can maximally explain the values of their child nodes at each particular sample size.

[44] In general, if one only wishes to use a Bayesian network to compute the results of conditional or marginal queries, then the directions of the given links are not as important as the overall link topology (Koller and Friedman, 2009). However, for distinguishing predictive versus diagnostic reasoning, or for examining the potential results of interventions versus observations, the link directionality will also be a critical factor in the model.

3.6 Application: Score-Based Structure Estimation of Brand Concept Maps for Retailers *Old Navy* and *Abercrombie & Fitch*

Analysis method. In order to illustrate the use of score-based Bayesian network structural estimation procedures, these methodologies were applied to the derivation of brand concept maps for two well-known retailers: *Old Navy* and *Abercrombie & Fitch*. Initially, a set of open-ended questions concerning these retailers was administered to a cross-section of undergraduates at a large Southwestern university. The responses to these open-ended questions were then analyzed by several team members and those recovered concepts which were repeated across multiple subjects were retained for further study. A pilot questionnaire was created based on these retained associations and was once again administered to the same set of undergraduates. The twenty-five brand associations which received the highest number of responses on this pilot questionnaire were then utilized in a larger-scale follow-up survey in which 800 undergraduates were asked to separately rate how strongly they associated each of these attributes with the retailer *Abercrombie & Fitch* and with the retailer *Old Navy*. The order in which these two retailers appeared in the questionnaires was reversed on half of the administered surveys. Responses were given on a simple five point Likert scale (anchored by 1 = *not at all* and 5 = *strongly associated*). Due to a small number of non-responses, the final number of recovered surveys differed slightly from 800.

Since the number of possible directed acyclic graph structures on *n* variables is super-exponential in *n* (Friedman and Koller, 2003), it is not necessarily possible to reliably estimate a full directed structure on all twenty five variables based on the number of data observations that were collected for this study.[45] Hence a representative subset of thirteen variables was selected from this larger set of brand associations.

In order to derive the structure of the corresponding brand concept maps based on these particular variables, a Bayesian search procedure was utilized, as implemented in the GeNie program suite, version 2.0 (available from genie.sis.pitt.edu). This algorithm begins with a random directed acyclic graph structure connecting all variables in the domain, and uses a greedy hill climbing procedure with random restarts to traverse the space of possible directed acyclic graphs based on those variables (Heckerman, 1995; Chickering, Geiger, and Heckerman, 1995; Chickering, 1996). At each stage of the search procedure, the algorithm examines every directed acyclic graph structure which is a neighbor of the current DAG (within the space of possible DAG structures), and iteratively moves to that neighboring structure which has the highest Bayesian structural score.[46] Within this search procedure, a 'neighboring structure' is one which is obtainable from the current structure by applying any one of three allowed structural perturbations: the addition of a directed edge, the deletion of a directed edge[47], or the reversal of the direction of an edge that is already included in the structure (Chickering, 1995, 1996; Koller and Friedman, 2009).[48] The Bayesian search procedure utilized here

[45] Since we further partitioned the data for each retailer by respondent gender, this resulted in approximately 400 data observations for each of the four data subsets (two retailers by two genders). Also note that the size of the space of directed acyclic graphs on 25 variables means that heuristic search methods will likely be required to search this space no matter what the size of the data set might be.

[46] Depending on the number of variables involved, a reasonable limit is sometimes placed on the number of allowable parents that each node may have.

[47] Edge deletions can potentially raise the overall structural score of the network if the removed edge provides redundant information about the states of the child node that it points to.

[48] Technically, one could accomplish an edge reversal by first deleting that edge and then adding back an edge with the reverse orientation at that same location, which would seem to make edge reversal a

also incorporates a particular prior distribution into the Bayesian score measure, namely the BDeu prior, which has several desirable properties, the most notable of which is that it assigns equal scores to all Markov-equivalent Bayesian network structures (Buntine, 1991; Heckerman et al., 1995).

Utilizing these techniques, the search for a Bayesian network based brand concept map was performed separately on four different subsets of the collected data: *Abercrombie & Fitch* - female respondents (henceforth termed “A&F-females”), *Abercrombie & Fitch* - male respondents (termed “A&F-males”), *Old Navy* - female respondents, and *Old Navy* - male respondents. Recall that the Bayesian search procedure operates via a greedy hill-climbing algorithm with random restarts within the large space of all directed acyclic graphs formed from the variables in the study. Since this implies that the algorithm approaches a maximum from multiple random directions, we took even further advantage of this randomization aspect by executing the algorithm multiple times within each of the four data sets until each data set yielded multiple runs in which the same maximum-scoring structure was returned. Such a result implies that the same maximal structure had been found beginning from multiple, highly different directions. Hence these multiple runs of the program which resulted in the same structure provided further evidence that the recovered structures were highly unlikely to simply

superfluous rule. However, removal of an edge often results in a lower Bayesian structural score (especially if the parent node of that edge has any appreciable explanatory power over the child node of that edge, above and beyond the influence that other possible collections of parent nodes have on that child node). Hence a strictly monotonic greedy hill-climbing procedure would not locate such a neighboring structure unless edge reversal is included as an explicitly permissible transformational rule. (See Chickering, 1995, 1996, and Koller and Friedman, 2009, for more details.)

represent local maxima, but rather, that they truly represented the global maxima for the Bayesian score as applied to that data set.[49]

Results and discussion. The brand concept maps for Abercrombie & Fitch and Old Navy which are derived by this procedure are presented in Figures 26 and 27 respectively. A comparative analysis of these four brand concept maps follows below.

Gender differences in the perception of trendiness. One critical difference that is readily apparent between males and females across both retailers is that females seem to regard "Trendy" as the central feature of their brand concepts, while males place "Trendy" in a much more peripheral position in both cases. Probing these differences even further, one finds that males seem to have a relatively superficial concept of "Trendy", since in each case they conceive of "Trendy" as being predictive of only some fairly overt brand characteristics. For instance, in the case of A&F, males view "Trendy" as predictive of brand characteristics such as "Bright Colors", "Crowded", and "Stylish" (which, for males' perceptions of A&F, is a maximally peripheral association). Similarly, males perceive of trendiness for Old Navy as merely being predictive of "Preppy", which in this case may be regarded as essentially synonymous with "Trendy", i.e., males seem to regard trendiness for Old Navy as simply indicating the form of trendiness that they perceive to exist at that retailer, namely "Preppy".

[49] Due to the heuristic nature of this approach (along with the exceedingly large size of the space of all possible DAG structures formed from these variables), the procedure may get "trapped" within a local "plateau" of structures which have the same Bayesian structural score (such as a set of members of the same Markov equivalence class, for instance). Of course, this possibility is strongly ameliorated by the algorithm's use of random restarts within each run. Nevertheless, some runs of the program within each data set would return idiosyncratic structures which had lower overall scores than the eventual structure which was accepted as the optimal one for that data set. For each of the four data sets, the structure that was declared to be the optimal one was returned the most often by the multiple runs of the program, and had the highest score out of all of the returned structures for that data set.

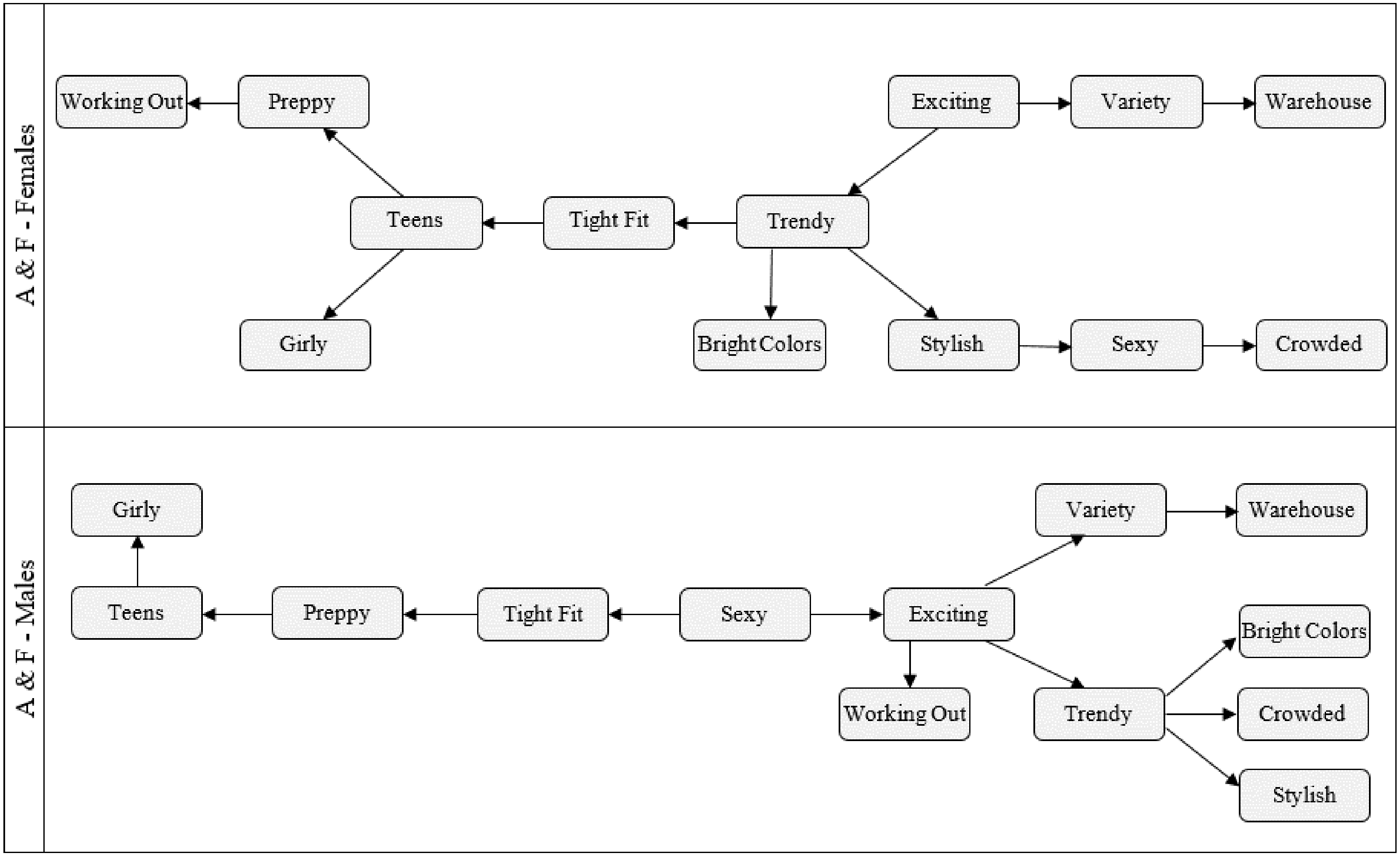


Figure 26. Brand concept maps for Abercrombie & Fitch

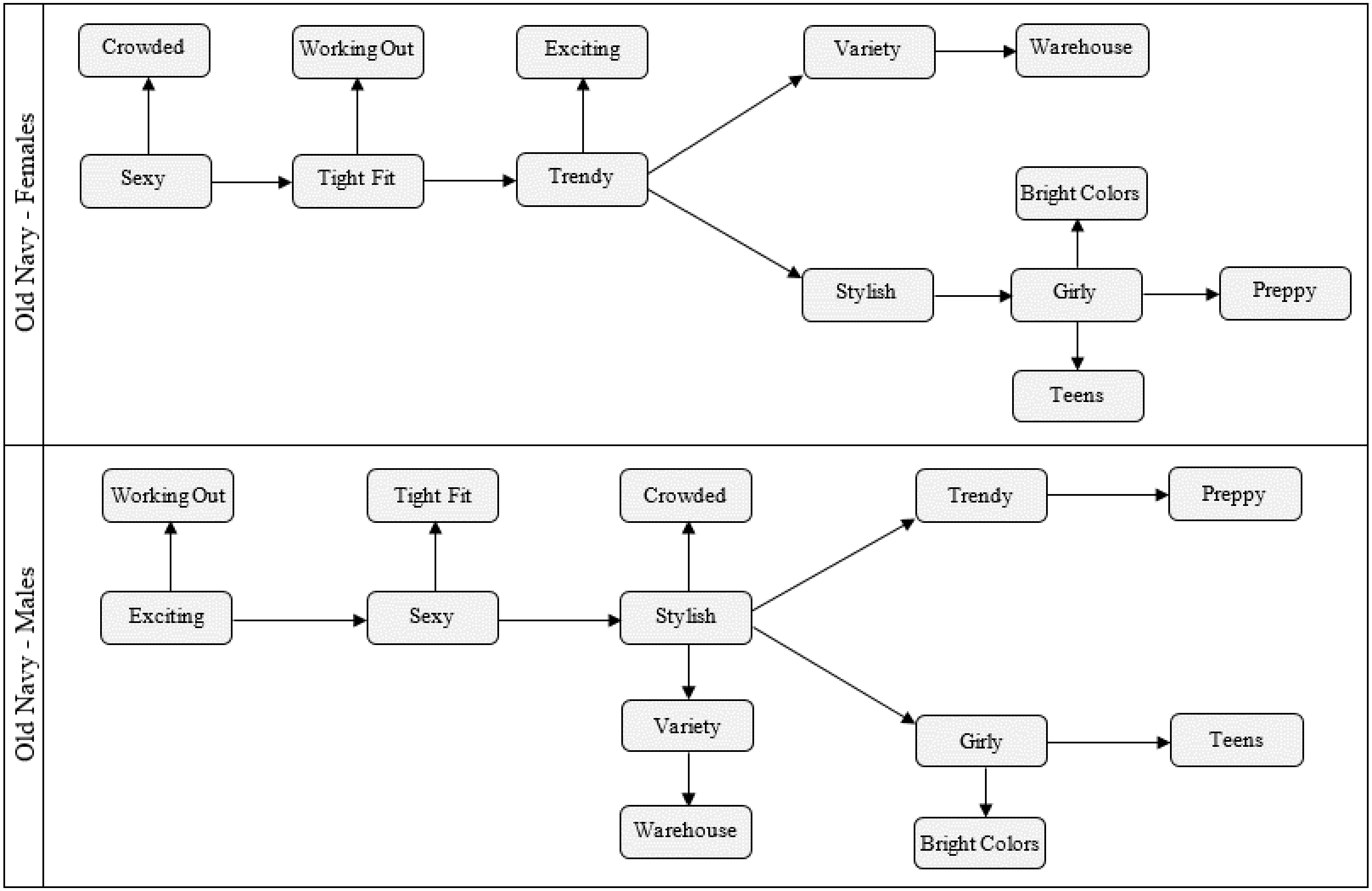


Figure 27. Brand concept maps for Old Navy

While males in both cases seem to regard “Trendy” as nothing more than a convenient ‘label’ for several other overt brand characteristics, females in both cases seem to have a much more nuanced interpretation of trendiness. For instance, within their brand concept maps for Old Navy, females see trendiness as being predictive of both stylishness and variety, as well as having an ‘excitement’ component. Further, within the brand concept map for Abercrombie & Fitch, females see trendiness as being diagnostic of “Exciting”, and predictive of “Stylish” (and hence also of “Sexy”, which is a child node of “Stylish”). However, within this same brand image, females also see “Trendy” as having a more superficial component, leading to “Tight Fit”, “Bright Colors”, as well as the entire cluster of ‘teenage’ concerns (such as “Teens”, “Preppy”, and “Girly”). Hence females see trendiness for Abercrombie & Fitch as having both a ‘serious’ component (“Exciting”, “Stylish”, “Sexy”) which is likely related to the deeper meaning of this concept for females, as well as a more trivial component (“Tight Fit”, “Bright Colors”, “Teens”, “Preppy”, and “Girly”) which likely describes what females perceive to be a component of the *type* of trendiness offered by this particular retailer.

Of course, such gender-based differences in the perceptions of trendiness should come as no surprise since clothing trends are known to form a system of non-verbal communication (Holman, 1980; McCracken and Roth, 1989), and further, as shown by Myers-Levy and Maheswaran (1991) and Myers-Levy and Sternthal (1991), there are systematic differences in males’ and females’ modes of information processing as well as their utilization of message cues. Hence, the finding that females seem to utilize the characteristics of both Abercrombie & Fitch and Old Navy to form more complex social codes involving responsiveness to fashion trends than do males is certainly to be expected based on former research. However, the directional brand concept maps derived via a Bayesian network analysis of this brand data also reveal some surprising differences even within females' views of the trendiness construct, as is detailed in the next subsection.

Excitingness and the trendiness/stylishness construct. As another indicator of the comparative complexity of females' perceptions of trendiness, the variables "Trendy" and "Stylish" have the same bivariate relationship within females' views of both retailers' brand images (namely "Trendy" directly implies "Stylish" in both cases), but the location of these two variables in reference to the remainder of the brand image differs in each case. For instance, within the A&F brand image for females, "Stylish" leads to "Sexy" and "Crowded", whereas "Trendy" leads to both "Bright Colors" and "Tight Fit", and hence indirectly to "Teens", "Preppy", and "Girly". Thus within the A&F image for females, it seems that the form of *stylishness* offered by the brand is a 'sexy' form of style, but the types of *trends* satisfied by the brand are trends toward tight-fitting clothes that are predictive of teens, preppiness, and girliness. On the other hand, within females' brand image for Old Navy, the form of *style* that females perceive the brand to offer is one of girliness, teens, preppiness, and bright colors, while the aspects of *trendiness* that females perceive the brand to offer comprise associations such as "Exciting" and "Variety". Hence, females seem to perceive Old Navy as catering to a 'teen' type of style, and a form of trendiness that seems more reminiscent of the type of shopping experience offered by the store ("Variety" and "Exciting") rather than a specific type of clothing style per se.

Now consider the fact that for the A&F-females brand image, "Exciting" leads to "Variety" and "Trendy". This indicates that females may perceive of A&F as being exciting because it offers variety and a specific form of trendiness which consists of a 'teen popularity' component as well as a sexy style component. In comparison, females seem to perceive of Old Navy as being trendy because it is an exciting and variety-laden place to shop. In other words, it seems as though females' view of Old Navy centers more around the shopping experience as being a reason for its trendiness (while its style is centered on teen concerns), while on the other hand females' view A&F as being exciting because of the trends it satisfies (which has a teen component as well as a sexy style component). Thus the directional relation between 'excitingness' and the general concepts of trendiness and style are effectively reversed for these two retailers, and this is an insight that is made apparent by the *directional* nature of the Bayesian network-derived brand concept maps.

Gender-based perceptions of "Working Out" and "Sexy". Although there are a myriad of differences between females' and males' brand concept maps for these two retailers, the differences between males' and females' perceptions of the roles of the "Working Out" variable and the "Sexy" variable are particularly instructive. For example, whereas males consistently perceive of "Working Out" as an outcome of "Exciting" across both retailers' brand constructs, females place "Working Out" in different positions in each case. Within the brand concept map for Abercrombie & Fitch, females place “Working Out” as a consequence of “Preppy”, while within the Old Navy brand concept map, females place “Working Out” as a direct consequence of “Tight Fit”.

These findings make sense in light of the directional differences in females’ perceived relationships between the excitingness construct and the trendiness/stylishness construct. As was described in the previous subsection, females seem to view the trendiness of Old Navy as having a major component that is simply based on the shopping experience at that retailer, i.e., the trendiness of the brand is, to a large extent, manifested as ‘excitingness’ and ‘variety’, both of which would describe the shopping experience rather than the clothes per se. Hence it makes sense that for this retailer, females see “Working Out” in utilitarian terms, i.e., as simply an outcome of “Tight Fit” (which is a logical relationship between these constructs) rather than as possessing any sort of stylistic relevance or expressiveness for the brand, since the brand is seemingly more characterized by the shopping experience than by the clothing styles.

On the other hand, as we also saw in the previous subsection, females tend to view the Abercrombie & Fitch brand as being exciting because it is trendy and has a sexy style, i.e., the A&F brand seems to predominantly be viewed via the stylistic statement it makes rather than by the shopping experience per se. Hence, in this case, the position in which females place “Working Out” within the brand concept map also makes a great deal of sense, because it is positioned as an outcome (or diagnostic indicator) of a stylistic component of the brand (namely “Preppy”).

The two genders also differ greatly in their placement of the brand association “Sexy”. Specifically, whereas females tend to place this construct in relatively peripheral positions, males on the other hand (perhaps not surprisingly) place “Sexy” very close to the center of their brand concept maps in each case examined. In fact, within their brand concept maps of Abercrombie & Fitch, males actually place “Sexy” as a central ‘source node’ (i.e., an ‘exogenous’ node which ‘drives’ all of its neighboring nodes, which in this case were ‘Exciting’ and ‘Tight Fit’). Furthermore, within their brand concepts of Old Navy, males place “Sexy” as a central node which serves to connect “Exciting” and “Stylish”.

Lastly, note that there are interesting gender-based differences in the perceptions of the role of the variable “Tight Fit” within these retailers’ brand concept maps, and these differences tend to parallel the findings made for the brand association “Sexy”. For instance, for female respondents, “Tight Fit” was either an immediate antecedent or an immediate consequent of “Trendy”, indicating that females viewed “Tight Fit” as a form of either predictive or diagnostic indicator of the overall trendiness of the brand. However males on the other hand consistently placed “Tight Fit” as an outcome of “Sexy”, which indicates that they are viewing this variable in much more direct physical terms than do the female respondents. Of course, these findings once again reinforce the notion that males and females are utilizing these two brands as different forms of consumption or style codes (e.g., McCracken and Roth, 1989), with females holding much more complex definitions of this code than do males.

4. Inference in Bayesian Network Structures

4.1 A Bayesian Network-based Measure of Brand Association Strength

Due to the directed nature of the links in a Bayesian network, the definition of connection strength measurements can be more complex than for comparable undirected networks (Jitnah, 1999). In their extensive study of such measures, Nicholson and Jitnah (1998) determined that the mutual information measure of dependence (Shannon and Weaver, 1949) has superior properties than do other forms of dependence measure when used as the foundation for a link strength calculus within Bayesian network structures.

In order to describe the development of such mutual information-based link strength measures for Bayesian networks, recall that the uncertainty of a random variable *X* can be quantified by its entropy *H(X)*, defined as follows (Shannon and Weaver, 1949):

$$H(X) \;=\; \sum_{x_i} P(x_i)\, log_2 \frac{1}{P(x_i)} \;=\; -\sum_{x_i} P(x_i)\, log_2 P(x_i) \qquad (31)$$

Utilizing the entropy measure of uncertainty, one can then define the mutual information between two random variables *X* and *Y* as the difference between the uncertainty of either of the two variables versus that variable's uncertainty given that the value of the other variable of the pair is known, as follows (MacKay, 2003; Ebert-Uphoff, 2007):

$$MI(X,Y) \;=\; H(Y) - H(Y|X) \;=\; \sum_{x,y} P(x,y)\, log_2 \frac{P(x,y)}{P(x)P(y)} \qquad (32)$$

where the entropy $H(Y|X)$ of the conditional distribution of *Y* given *X* is defined as the weighted average of the uncertainty of *Y* given each possible value of the conditioning variable *X*, viz., $H(Y|X) = \sum_{x_i} P(x_i)\, H(Y|x_i)$. Furthermore, note that since the joint probability $P(x,y)$ in Equation (32) can be written as the product of the associated conditional and marginal probabilities, i.e., $P(x,y) = P(x) \cdot P(y|x) = P(y) \cdot P(x|y)$, we can also express the mutual information between *X* and *Y* in the form:

$$MI(X,Y) \;=\; \sum_{y} P(y) \sum_{x} P(x|y)\, log_2 \frac{P(x|y)}{P(x)} \qquad (33)$$

Since Bayesian network structures involve graphical representations of conditional independence relations, we must extend the mutual information measure to such conditional distributions. For example, the mutual information between variables *X* and *Y* conditional on knowledge of a third variable *Z* can be defined analogously to the unconditional case (MacKay, 2003; Ebert-Uphoff, 2007), namely as $MI(X,Y \mid Z) \;= H(Y|Z) - H(Y|X,Z)$, where the conditional entropy is given by $H(Y|X,Z) \;= \sum_{x,z} P(x,z)\, H(Y|x,z)$, i.e., by the weighted average of the conditional entropy of *Y* given knowledge of the states of *X* and *Z* (where the weighted averaging is performed over all possible configurations of the states of *X* and *Z*). Expansion of this expression yields a conditional mutual information expression which is analogous to the unconditional version given in Equation (32), as follows (MacKay, 2003; Ebert-Uphoff, 2007):

$$MI(X, Y \mid Z) = \sum_{x,z} P(x,z) \sum_{y} P(y|x,z) \log_2 \frac{P(y|x,z)}{P(y|z)} \tag{34}$$

In order to focus this measure on the quantification of the strength of one particular link connecting variable X to variable Y, we follow Ebert-Uphoff (2007) and define the conditioning set Z in Equation 34 as being the set of all *other* graphical parents of Y (i.e., those graphical parents of Y which are disjoint from X) within the given Bayesian network structure. In this manner, the act of conditioning on all other parents Z of Y which are disjoint from the other parent X renders the link $X \rightarrow Y$ the only active predictive link into variable Y, and hence the recovered strength measure will uniquely correspond to just this one link. Hence when the conditioning set Z in Equation 34 corresponds to the set of parents of Y which are disjoint from X, we refer to this measure as the link strength $S(X \rightarrow Y)$.

Based on the meaning of the mutual information measure, the strength $S(X \rightarrow Y)$ of a given link $X \rightarrow Y$ as defined by Equation 34 represents the weighted average reduction in the uncertainty about the value of Y which can be attributed to knowledge of the state of the parent node X, given each possible combination Z of the other parent nodes of Y (Ebert-Uphoff, 2007). This value can also be thought of as a measure of the amount of information contained in that link, and also as a relative measure of the degree to which the distribution of the antecedent variable X will affect the distribution of the consequent variable Y (Nicholson and Jitnah, 1998; Jitnah, 1999).

4.2 Application: Link Strength Analysis of the Taco Prima Brand Concept Map

As an illustration of its usefulness in quantifying the relationships within brand concept maps, the link strength measure $S(X \rightarrow Y)$ defined by Equation 34 was applied to the Taco Prima brand concept map derived in Section 3.3. In each case, the link strength values $S(X \rightarrow Y)$ were computed via a large spreadsheet analysis in which the values of each parent of the target variable Y were conditioned on prior to computing the strength of the connection from X to Y in order to ensure that the link strength measure was specific to just the one directed link $X \rightarrow Y$. The resulting link strength values for the Taco Prima brand concept map are shown in Figure 28 (in which we utilized the one-letter abbreviations for each brand association in order to facilitate displaying the link strength values directly within the same diagram as the association names).[50]

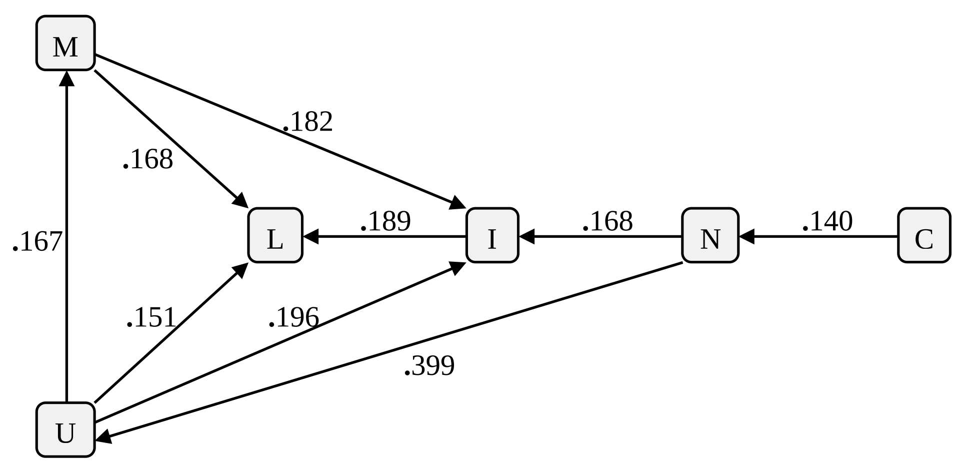


Figure 28. Link strength analysis for the Taco Prima brand concept map

[50] Recall that association "C" stands for "Inexpensive" (i.e., "Cheap"), while "I" stands for "Innovative".

Clearly, the one link strength value that stands out in this brand concept map is that of the link from "New Things" to "Unique". In fact, $S(NewThings \rightarrow Unique)$ is over twice as large as any other link strength quantity within this brand concept map. From a strategic viewpoint, this finding is intriguing because this particular extremely strong link is located precisely in the center of the critical pathway $Inexpensive \rightarrow NewThings \rightarrow Unique \rightarrow Memories$, which is the pathway that was hypothesized earlier to be the core driver of brand equity for this firm. Furthermore, this critical pathway is the one which was identified (through Markov equivalence class analysis) to possess all of the allowable cognitive variations which are still consistent with the dependence and independence relations in the collected data.

This finding raises the interesting conjecture of whether the permissible cognitive variations[51] which can exist for a given brand concept structure essentially *must* encompass any extremely strong links in that structure. Such a conjecture may hold true since such extremely strong links possess more information regarding the variables involved than do weaker links within that structure, and if consumers are going to hold differing directional (i.e., predictive and diagnostic) beliefs about a brand's associations, it is quite plausible that these differences would center around variables which hold a great deal of information for those consumers' understanding of the brand associations involved.

[51] By a permissible cognitive variation, we mean a structural variant which neither contradicts any established conditional independence relations nor establishes any new ones that were not identified in the data.

Stated more simply, if consumers are going to differ over something, they are likely to differ over something that has a lot of meaning for them. In fact, a converse conjecture would be that quite possibly, a particular link (such as the link between *New Things* and *Unique* in the Taco Prima brand concept map) is quite important (i.e., has a high link strength value, and hence holds a great deal of information) *because* it serves to differentiate between different consumers' cognitions about the brand. Furthermore, this converse conjecture may be especially true when the inter-group or inter-consumer differences are directional in nature (as they are here), since directional beliefs are likely to possess an "if-then" character, and such "if-then" beliefs can serve as informal hypotheses of what causes or leads to what within a brand concept or a more general consumption domain (Hoch and Deighton, 1989; Sirsi, Ward, and Reingen, 1996).

4.3 Application: Link Strength Analysis of the Old Navy and Abercrombie & Fitch Brand Concept Maps

Link strength analysis was also applied to the brand concept maps for the retailers Abercrombie & Fitch and Old Navy. Once again, the conditional mutual information-based link strength measure defined in Equation 34 was utilized. The resulting link strength values for both the males' and females' brand concept maps for each of these retailers are shown in Figures 29 and 30. Since these are more extensive brand concept maps, the resulting link strength patterns show a large degree of variability across the four retailer brand concepts studied, but some interesting patterns do emerge.

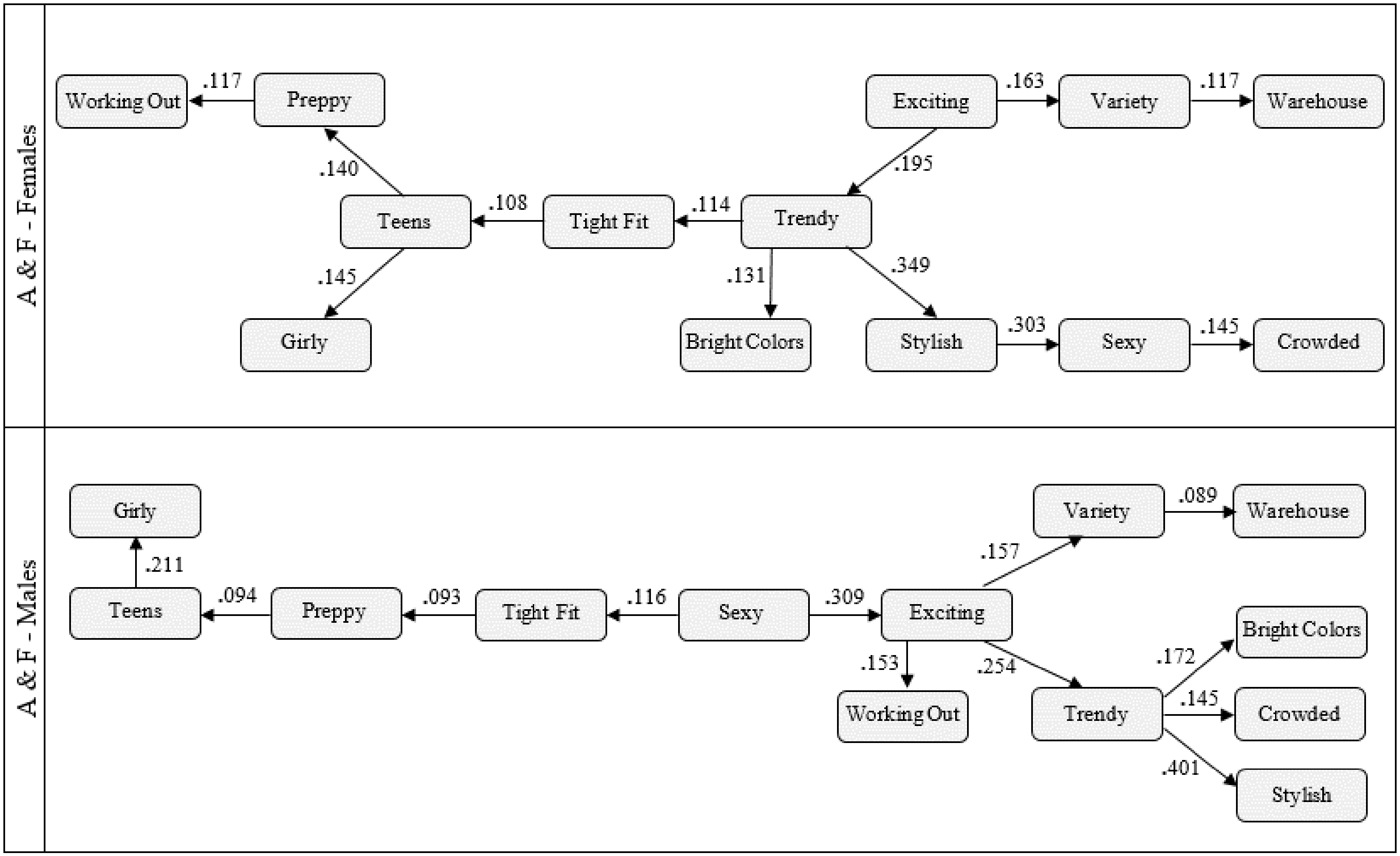


Figure 29. Brand concept map link strength analysis for Abercrombie & Fitch

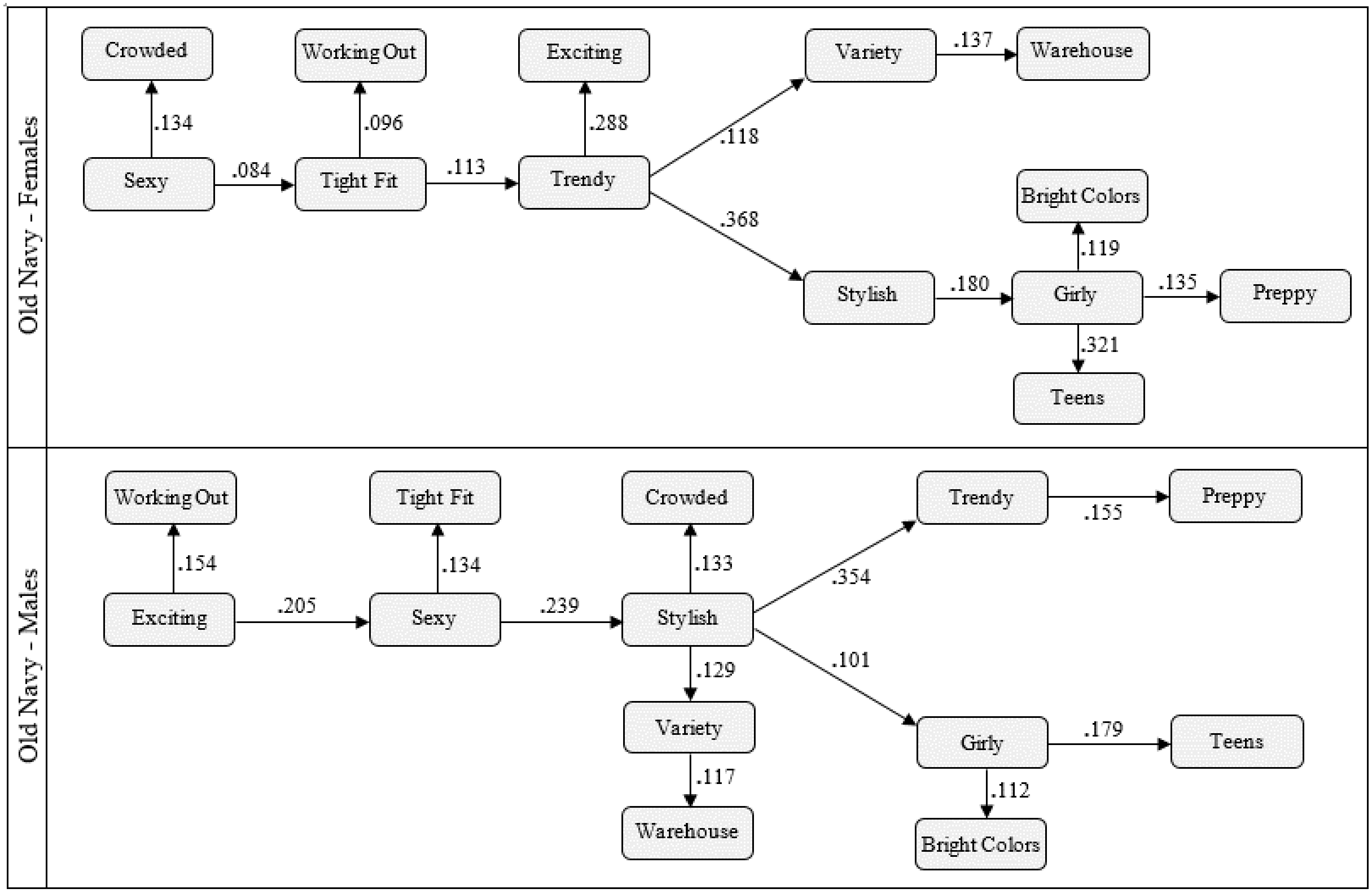


Figure 30. Brand concept map link strength analysis for Old Navy

Since we have four brand concept maps to compare, we can address questions such as whether the more 'central' links between brand associations tend to be stronger than the more peripheral links. In this case, the answer quite obviously is "no", since there are many peripheral links which have very high link strength scores, while there are a multitude of links near the center of each network which are comparatively very weak. Hence, at least based on this data, it does *not* seem that more centrally located links will necessarily be any stronger than more peripheral links within the same network.

A second major issue which we can address by having four comparable brand concept maps is the determination of whether consumers' brand concept maps tend to exhibit high link strengths between brand associations that can be regarded as partially synonymous with one another. Since we are not directly asking consumers to determine which brand associations have the same or similar meanings, the best way we can operationalize the concept of "synonymous" would be to utilize the notion that if one concept has a very similar or identical meaning as another, then each of those concepts can be used interchangeably across virtually all situations. In the case of a brand concept map, we can conceptualize "using" a brand association as "applying" that brand association in a particular way, i.e., as having that brand association either predictively or diagnostically imply another particular brand association across multiple contexts. Examining the four brand concept maps shown in Figures 29 and 30, there are three specific instances where consumers in all four maps consistently link the same two variables together in either a predictive or diagnostic relationship, and hence these variable pairs are likely to have similar operational meaning for the consumers involved.

One such instance of brand associations which directly entail one another across all four brand concept maps is the relationship *Variety* → *Warehouse*. Note that this relationship makes intuitive sense because consumers may tend to perceive of a large product assortment as being reminiscent of a 'discount' or 'warehouse' type of retailing environment. However, despite the fact that the brand association *Variety* is predictive of the brand association *Warehouse* in all four situations, the strength of the corresponding link *Variety* → *Warehouse* is actually quite low in all four brand concept maps, ranging in size from 0.089 to 0.137. Hence consumers in all four situations may see *Variety* as being predictive of a warehouse environment (and conversely perceive of a warehouse environment as being diagnostic of high variety), but despite the apparent ubiquity of this directional relationship, the data shows that this directional association is fairly modest in size.

Similarly, all four maps possess a direct link between the associations *Girly* and *Teens*,[52] and once again the fact that these two brand associations are so consistently linked makes intuitive sense. Interestingly, the directional relationship between these two brand associations has a somewhat "intermediate" strength across all four brand concept maps, with the link strength between *Girly* and *Teens* ranging from a low of 0.112 to a high of 0.321.

[52] The direction of the link between *Girly* and *Teens* is consistent within each retailer's brand concept maps, but is reversed between the two retailers. However, this just means that consumers in each case have opposing views of which of these two associations is predictive of the other versus which is diagnostic of the other. The point remains that in all four brand concept maps, these two brand associations have a direct implicational relationship between them.

Finally, note that the direct link between variables *Stylish* and *Trendy* is conserved across all four brand concept maps. Once again, the link directions are not consistent in all four maps, but all four maps do possess a direct connection between these two concepts. Interestingly, the strength of the inter-concept link between *Stylish* and *Trendy* is actually quite high in all cases, ranging between 0.349 and 0.401 across the four brand concept maps.

Similarly, there are cases where each gender is relatively consistent across brands, but where the two genders differ from one another within the same brand. For instance, consider that males consistently place a direct link between *Sexy* and *Exciting* and between *Sexy* and *Tight Fit*.[53] Interestingly, in both cases the link from *Sexy* to *Exciting* is quite a bit stronger than the link from *Sexy* to *Tight Fit*. Now consider the fact that both *Sexy* and *Exciting* are abstract or 'interpretive' descriptors, whereas *Tight Fit* is a 'concrete' or 'directly observable' characteristic. Hence, in this case, the link that males perceive between two directly connected interpretive descriptors is stronger than the link that they perceive between an interpretive descriptor and a directly observable characteristic.

This, of course, raises the question of whether the difference in link strengths between descriptor types which was noticed for certain characteristics of males' brand concept maps is a more general phenomenon. As a partial answer to this question, consider that for females' brand concept maps, the brand associations *Tight Fit* and

---

[53] The predictive versus diagnostic direction for the link that males perceive between *Sexy* and *Exciting* is reversed between each retailer, but males do perceive a direct link between these two associations across both retailers.

Exciting are linked by *Trendy* (rather than by *Sexy* as was the case in males' brand concept maps). What is fascinating here is that once again we see that the link between the more abstract or interpretive descriptors (namely *Trendy* and *Exciting*) is quite a bit stronger than the link between an abstract descriptor (*Trendy*) and a directly observable characteristic (*Tight Fit*), and furthermore, this pattern is consistent across both retailers' brand concept maps.

With this observation in hand, we can come back and re-examine the earlier finding that there was an appreciable link strength difference between the three cases in which we had specific links that were conserved across all four brand concept maps (i.e., the three cases which were hypothesized to represent approximate synonymy relationships within the context of these brand concept maps). In the first of these three cases, we saw that there was a relatively weak link strength between *Variety* and *Warehouse* across all four brand concept maps. Interestingly, both of these descriptors are fairly 'concrete' in nature, i.e., even though there may be differences across individuals in terms of how they view the relative levels of variety and 'warehouse feel' that exist at specific retailers, both of these descriptors do refer to characteristics that are externally observable. Hence we have that the *de facto* synonymy relationship between two externally observable descriptors has a relatively low link strength, which is an observation that does fit in with the previously stated hypothesis concerning the differing link strengths that exist across descriptor types.

Now consider the synonymy relation that was found to be intermediate in strength, namely the relationship between *Girly* and *Teens*. In this case, it is somewhat harder to declaratively classify these two descriptors, but clearly one can observe whether a retailer is frequented by teens (or caters to teen fashions), while the descriptor 'Girly' is more of an interpretive characteristic rather than something that one can 'point to' in a physical sense.[54] Hence, in this particular example, we have that a relation between an abstract or interpretive descriptor of the brand and a more externally observable characteristic of that brand has an intermediate level of strength, which is again consistent with the earlier stated hypothesis.

Finally, consider the potential synonymy relation between *Stylish* and *Trendy* that was discussed earlier. In this case, both descriptors are clearly interpretive in nature, and it is also the case that the link between these interpretive descriptors is very strong across all four brand concept maps. Therefore, this observation once again follows from the earlier stated hypothesis concerning the pattern of link strength differences that occur across differing descriptor or association types.

One can also hypothesize about the potential reasons for this observed pattern of link strength differences that seems to exist between interpretive descriptors, between concrete (observable) descriptors, or between these two classes of descriptor. Clearly, the most immediate potential explanation for these differences resides in the 'fan effect' which is hypothesized to underlie directional relations within the associative networks literature. It seems reasonable to assume that because abstract or interpretive descriptors can potentially be applied to many concrete or observable situations, that such

[54] Of course, for a retailer that specifically caters to females, once could 'point to' the 'Girly' nature of that retailer's brand image. However, both Abercrombie & Fitch and Old Navy are retailers which cater to both genders, and hence the degree of 'girliness' of these retailers would likely be more of an interpretive judgment rather than an observable description.

interpretive descriptors will have a greater number of substructural associative links fanning out from them, and hence there is a greater chance that the fan structures from two such interpretive descriptors will meet, and also it is more likely that if they do meet, that they will meet in multiple places. On the other hand, a concrete or directly observable descriptor may be too 'tied' to a specific context to have an appreciable number of links fanning out from it, and hence it is far less likely that the fans emanating from two such concrete descriptors will meet each other in very many places. Finally, in the case of an interpretive descriptor and an observable descriptor, the fact that the fan structure associated with the interpretive descriptor will be relatively large and that of the observable descriptor will be comparatively limited would seem to imply that one will observe a level of link strength that is intermediate between that of two concrete or observable descriptors and that of two abstract or interpretive descriptors.[55]

4.4 Informational Propagation in Bayesian Networks

As described in Section 2.2 of this thesis, the Bayesian network representation of a brand concept map exploits the conditional independence relations that exist in the data in order to break down the complex global factors that constitute the joint distribution of the brand association variables into simple local factors that each typically involve just a few variables. In this manner, the Bayesian network representation of a brand concept map provides a very compact representation of the overall joint distribution of the full set

[55] One should note that despite the discussion of associative mechanisms, the links being referred to here are decidedly directional in nature. As described in Section 1.4, the associative fan effect is a widely hypothesized mechanism by which an underlying associative stratum can result in a set of directional relations among attributes.

of variables involved in that brand concept (Pearl, 1988). In fact, each node of a Bayesian network can actually be regarded as a local distribution of the values of the variable associated with that node, conditional on each possible state of its ‘graphical parent’ nodes (i.e., conditional on the values of its immediate predecessors in the directional structure of that network representation).[56]

Since the Bayesian network is simply a graphical representation of the overall multivariate joint distribution of the variables that constitute the brand concept map, the answers to specific propositional queries can be computed by using the mathematical properties of this distributional representation. For instance, one may wish to know how observing or fixing the value of one or more variables in the network will affect the distribution of values within the remaining network variables. For example, considering the *Abercrombie & Fitch* or *Old Navy* brand concept maps shown in Figures 27 and 28, a manager might be interested in ascertaining the likely effect that increasing the level of ‘bright colors’ within his or her store might have on the distributions of the remaining network variables. Such propositional queries can be answered in a concise and tractable way by exploiting the factored representation of the joint distribution that is represented by the corresponding Bayesian network model of the data.

In essence, there are two categories of network structure to consider here: structures in which there is just a single pathway between any two network variables, and structures in which there are multiple pathways between variables. These structures are

[56] These distribution parameters can typically be estimated by a maximum likelihood calculation, again making use of the global and local decomposition properties afforded by the Bayesian network representation of the data (Pearl, 1988; Koller and Friedman, 2009). This is especially straightforward in the case of multinomial data such as the type often encountered in typical Likert-style marketing surveys.

generally known as *singly-connected networks* and *multiply-connected networks* respectively (Pearl, 1988). In the case of singly-connected networks, the pioneering effort to derive a tractable and precise informational propagation algorithm which can consider both the predictive influences of those directional links coming into each node as well as the diagnostic influences from directional links that are 'coming out of' that node was given by Kim and Pearl (1983). Since that time, several methods have been developed which reduce multiply-connected networks to the singly-connected case through various processes (Koller and Friedman, 2009).

The most commonly used such procedure for reducing a multiply-connected network to a singly-connected structure is that of *clique-tree propagation*, which was developed through the efforts of multiple researchers during the 1980's and 1990's (see, for instance, Huang and Darwiche, 1996, for a good introduction to this general class of procedures). In essence, this technique groups together those variables which share a common 'child' node (a process called *moralization*, since it amounts to joining or 'marrying' the parents of a common child node), and then triangulates each resulting cluster in order to form a clique structure (i.e., a subgraph in which each variable is directly connected to each other variable within that subgraph). In most networks, there are multiple possible ways to accomplish this structural transformation, and the method singles out those transformations which result in a *tree structure* connecting the cliques (i.e., there is branching, but no cycle or loop structures within the structure), and then further singles out those tree structures which obey a form of informational monotonicity generally known as the *running intersection property*, which requires that any pathway between two cliques must contain all variables that belong to the intersection of those two cliques.[57] Such a structure consisting of a tree connecting cliques and which satisfies the running intersection property is called a *clique tree*.[58]

Once this transformation of the original network into a clique tree structure is accomplished, each pair of neighboring cliques will have one or more variables in common, and the set of variables that are shared between two neighboring cliques is termed the *sepset* for that pair of cliques. Furthermore, since the clique tree construction obeys the running intersection property, one can show that each sepset forms a *minimal separator* within the original Bayesian network, i.e., a minimal set of variables which, when conditioned on (through intervention or observation), will render those nodes on either side of that separator conditionally independent of each other. The propagation of predictive and diagnostic influence among any two cliques can then be systematically accomplished by multiplying each probabilistic factor in the cliques on one side of the sepset which probabilistically separates those two cliques and then marginalizing out the variables from that side of the sepset which are not themselves present in that sepset (Huang and Darwiche, 1996; Koller and Friedman, 2009). Hence, the system of sepsets in the clique tree structure derived from the original network form a maximally efficient collection of variables to condition on in order to perform informational propagation.

[57] Among other things, this property essentially ensures that no information is lost when transmitting information between those cliques, and that information propagation through the derived structure will mimic that which would have been obtained via an exhaustive marginalization process in the full joint distribution (Huang and Darwiche, 1996; Koller and Friedman, 2009).

[58] We also assume that the scope each probabilistic factor in the underlying distribution is contained within a single clique (which is a property often known as *the family preservation property*). Again, see Huang and Darwiche (1996) and Koller and Friedman (2009) for more details.

4.5 Application: Information Propagation in the Taco Prima Network

We have applied the clique and sepset-based information propagation mechanism to the Bayesian network model of the Taco Prima brand concept map[59] which was derived via constraint-based methods in Section 3.3. The associated clique tree computations were performed with the GeNie software package (Druzdzel, 1999). The initial "baseline" (or "prior") distribution of the variables in the Taco Prima brand concept map is shown in Figure 31. [60]

To demonstrate the usefulness of the information propagation mechanism within this network, assume that management is considering a new promotional effort to strongly reinforce and further improve Taco Prima's inexpensive brand image, but they are unsure as to how effective this promotional strategy may be in fostering the ultimate goal of improving consumer affect (i.e., 'Memories') toward the brand. Since all of the variables in the Taco Prima network are interconnected and also have positive link strengths (Section 4.2) as well as strong, positive, and highly significant bivariate Pearson correlations, it might be natural for management to think that the higher they can drive consumers' perceptions that the brand is very inexpensive, the greater the corresponding ultimate effect on "Memories" (i.e., brand affect) is likely to be. However, what this supposition does not take into account is that there are multiple interacting pathways by which predictive and diagnostic influence can propagate within the network.

---

[59] Since each of the four brand concept maps for the retailers *Old Navy* and *Abercrombie & Fitch* are singly-connected tree structures, they have very simple information propagation properties, and hence are not as instructive as are the informational propagation properties of the Taco Prima network. Hence we chose to illustrate the information propagation method by using just the Taco Prima network.

[60] The abbreviations "v1" through "v5" in each distribution represent the five possible response values from the original questionnaire.

For instance, to attempt to answer management's query about the likely effects of an effort to dramatically increase consumers' perceptions that Taco Prima is inexpensive, we can enter specific evidence (such as *Inexpensive* = 5) into the network and then allow the information propagation algorithm trace its predictive and diagnostic influences throughout the network. In this way, one obtains an updated (or 'posterior') distribution which incorporates this evidence into the probabilities of the variables involved. The posterior distribution based on incorporation of the evidence value *Inexpensive* = 5 is shown in Figure 32.

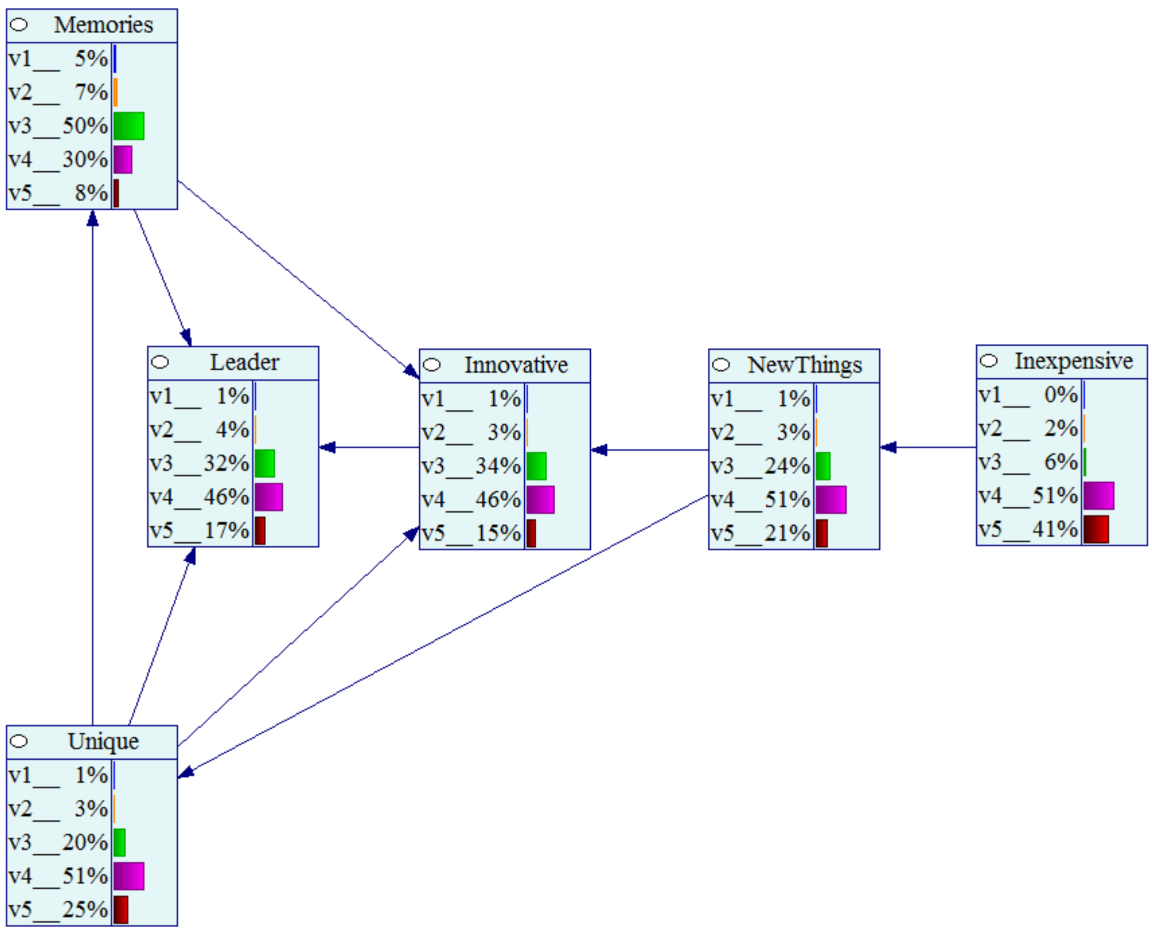


Figure 31. Prior distribution of variables in the Taco Prima brand concept map

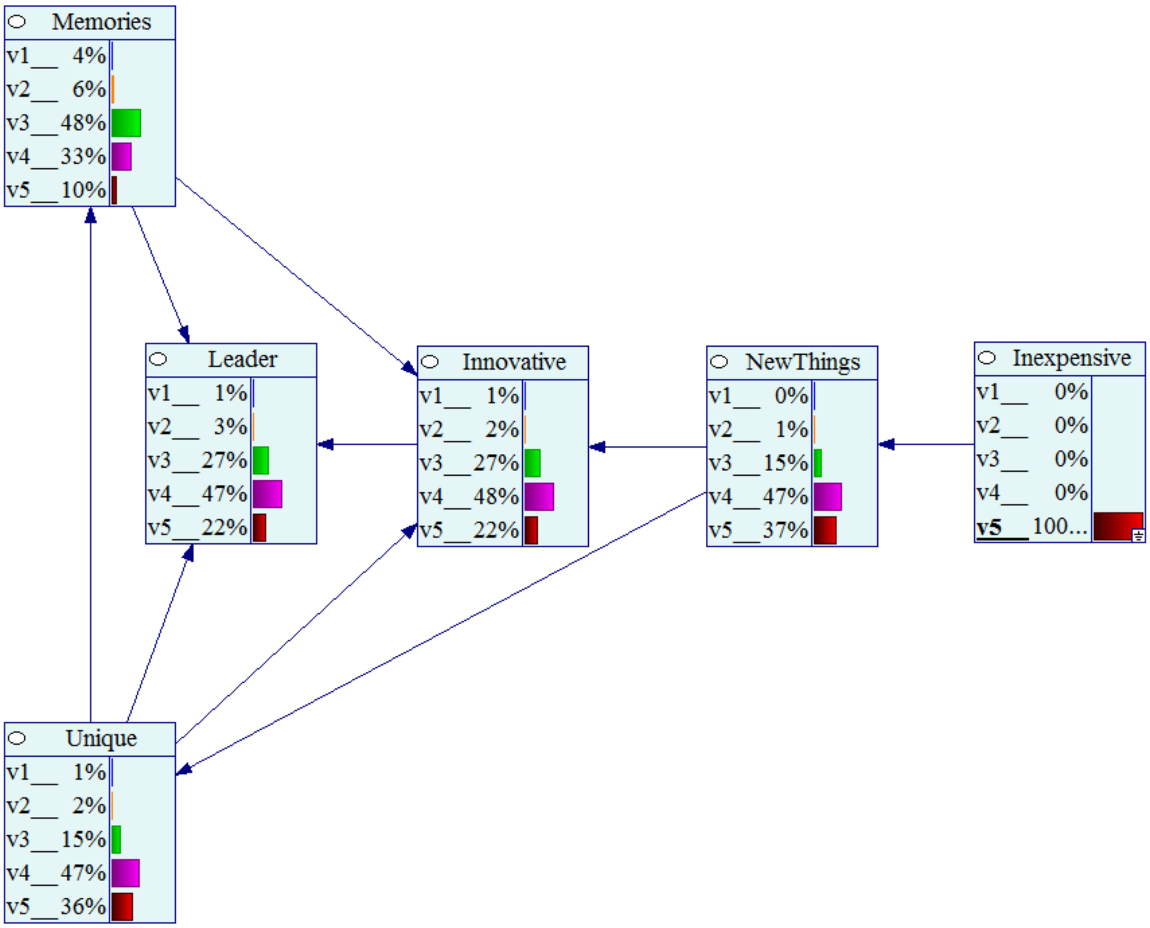


Figure 32. Posterior distribution of variables in the Taco Prima brand concept map after incorporation of evidence *Inexpensive* = 5.

Clearly, many of these individual posterior distributions do differ in expected ways from the prior distributions shown in Figure 31. For instance, as expected, evidence that *Inexpensive* = 5 leads to a decrease in the probability that either *New Things* or *Unique* will have values of 3 or 4, and an increase in the probability that *New Things* and *Unique* will each have a value of 5. Similarly, for *Innovative* and *Leader*, the posterior distributions for these variables have a reduced probability of the response scale value of 3 and an increased probability of having a value of 5, but in the case of these two variables, the probability of having a value of 4 is essentially unchanged from the prior distribution. (Hence *Innovative* and *Leader* show a qualitatively different pattern of changes between the prior and posterior distributions than do *New Things* and *Unique*.)

Now to address management's initial question, consider the posterior distribution of the variable *Memories*. Despite the strong positive correlations among all variables in this network, as well as the positive link strength values for every inter-association link, the posterior distribution of the variable *Memories* differs little (if at all) from its distribution prior to the incorporation of evidence about consumers' perceptions that the brand is inexpensive. In essence, contrary to some very reasonable-seeming assumptions, the interaction of predictive and diagnostic information flow within this network makes the proposed increase in consumers' perceptions of the brand as very inexpensive (i.e., *Inexpensive* = 5) essentially ineffective at appreciably altering the distribution of the variable *Memories*. Hence according to this brand concept network, it would likely *not* be effective for management to try to improve brand affect (i.e., 'Memories') via driving the image of the brand towards the extreme "inexpensive" end of the price perception spectrum.

Similar observations about the informational updating properties of this brand concept can be made by incorporating various facets of evidence about each of the other variables in the network. In each case, the procedure is essentially the same as that described above.

5. Conclusions and Extensions

5.1 General Summary and Conclusions

We initially set out to develop a means of analyzing the structure of consumers' brand concepts which would extend the pioneering approach of Roedder John et al. (2006). In keeping with the Brand Concept Mapping approach of Roedder John et al., we sought to uncover the specific link structure that serves to connect the brand associations so that it is clear which brand-related attributes are directly connected, and which are indirectly connected through one or more intermediaries. Further, we sought to follow up on the suggestion by Roedder John et al. (2006) that there is likely to be a directional component to the links that consumers perceive to exist between various brand-related attributes, and that the Brand Concept Mapping approach could be notably extended if not just the connection pattern, but also the specific form of the links themselves (i.e., their internal make-up or constituency) was explored.

Through an examination of the marketing literature, it became clear that the marketing domain is replete with examples in which such belief structures are indeed directional in nature, ranging from causal attribution at all levels (e.g., Folkes, 1988; Weiner, 2000), directional priming effects (Nedungadi, 1990; Holden and Lutz, 1992; Farquhar and Herr, 1993), and even simple "if-then" reasoning within consumers' conceptualizations of their environment (e.g., Hoch and Deighton, 1989; Sirsi, Ward, and Reingen, 1996). Even consumers' typical purposes in considering a purchase in the first place are known to be causal or directional in nature (Folkes, op. cit.).

Furthermore, by examining the core associative networks literature on which the brand concept map construct was originally based, one finds that researchers quite often utilize associative structures to explain or model directionally asymmetric phenomena. In fact, such directional associative phenomena have even been used to model various phenomena within the marketing domain itself (e.g., Ulhaque and Bahn, 1992; Lei, Dawar, and Lemmink, 2008). Thus we sought to extend the Brand Concept Mapping procedure through the explicit consideration of directional links among perceived brand associations.

A further analysis of the internal structure of directional links between attributes revealed that rather than simply being links which specify an antecedent and a consequent, directional associations also possess an internal structure that is sensitive to the difference between observational versus interventional conditions, and which clearly distinguishes between the predictive and diagnostic directions within directional relationships. Furthermore, it is known that individuals' reasoning patterns typically differ between the predictive and the diagnostic directions of the same directional or causal relationship (Tversky and Kahneman, 1980). Therefore, the fact that the techniques utilized here can discern not just the connectivity between brand attributes, but also the directionality of that relationship can be extremely useful for understanding the directional reasoning patterns typically held by consumers.

As shown earlier, at the level of two-variable structures (and assuming purely observational conditions), a directional relationship can appear and behave just like a non-directional or purely correlational relationship (since diagnostic reasoning will allow the consequent of the directional relationship to show some influence over the antecedent of that relationship). However, it is at the level of *three-variable* substructures where we begin to see an unavoidable differentiation between directional and non-directional relationships. In fact, it is through a closer examination of the differences among the three-variable substructures known as the causal chain (A → B → C), the diagnostic chain (A ← B ← C), the common cause structure (A ← B → C), and the common effect structure (A → B ← C) that we begin to see the importance of not just dependence versus independence, but additionally the *conditional independence* construct as a means of distinguishing among otherwise observationally or correlationally indistinguishable configurations.

As shown earlier, the limits of what the conditional independence construct can distinguish are known as Markov equivalence classes. Within a Markov equivalence class, each structure will have the same conditional independence and dependence properties, but structures in different Markov equivalence classes will have differing properties of conditional independence.

We then applied the principles of Markov equivalence and conditional independence to brand association data by utilizing both constraint-based and score-based structural discovery algorithms which take advantage of these conditional independence based differences among the various types of directed links and substructures which can connect the brand associations. The recovered structures based on these discovery methods typically possess specific regularity properties, and are often collectively known as Bayesian networks. A further analysis of the Markov equivalence classes (i.e., the observationally indistinguishable sets of directed structures) within the Bayesian network representation of a brand concept map led us to realize that this methodology can prescribe the specific set of directional cognitive variations that can co-exist within a specific brand concept structure and still be consistent with the conditional independence properties inherent in the data.

Lastly, we applied the specific structure of the directional links within a Bayesian network representation of consumers' brand association data in order to provide two additional computational tools. First of all, we provided a link strength measure which can be utilized to further probe the relationships among the variables in the domain, and we found that additionally, this measure seems to be sensitive to the abstract versus concrete nature of the brand attributes themselves. Secondly, we utilized the directional structure of the inter-attribute links to allow for the application of information propagation tools, which can be used to precisely compute the posterior distribution of the variables in the network based on any given set of observational queries or evidence presented to the network.

In conclusion, we feel that this effort has yielded a valid approach which successfully answers the call by Roedder John et al. (2006) for an extension to their pioneering Brand Concept Mapping approach, and which can explicitly consider directional relations among brand associations. Further, by pursuing the various ramifications of this approach within the marketing domain, we feel that valuable additional tools have been provided which can be used to explore aspects of consumption phenomena which had previously been relatively inaccessible to marketing researchers.

5.2 Hypotheses and Future Extensions

The techniques explored in this thesis reveal a set of concepts and a related set of tools which can potentially be utilized by marketing researchers to explore new and important phenomena within multiple marketing domains. However, these tools are not without their limitations. For example, all of the structural discovery methodologies utilized in this thesis still strongly depend on the validity and completeness of the initial round of brand association elicitation in order to have a valid corpus of brand characteristics to work with. Hence, just as is the case with other multivariate methodologies within the marketing domain, we still must depend on the invaluable contribution of exploratory and qualitative marketing research techniques.

In addition, in order for the various structural discovery methods utilized here to be guaranteed to converge to the single most optimal structure or set of structures (as opposed to merely a satisfactory or comparatively optimal such structure), one also needs to assume that all relevant causal variables or background factors have been included in the model. Just as is the case with other multivariate techniques, such an assumption is difficult or impossible to guarantee in practice, and hence one needs to regard the results of such studies with the same degree of caution as one does with the results of most other multivariate methodologies. In other words, the Bayesian network technique is not a panacea for an improperly specified model.

In addition, we note that the brand concept map structures that are derived from the Bayesian network structural discovery protocols do not explicitly show the links between the brand attributes and the core brand itself. Rather, it is implicitly assumed that all attributes that are related to the brand essentially belong to a general structure with the brand as its center. However, this is an area in which further research may be quite valuable, since it may be possible to include specific links to the brand itself by modifying these techniques appropriately.

Finally, it must be pointed out that by using these particular structural discovery protocols, we are not asking directly about whether a link exists between brand association A and brand association B. Rather, we are simply asking consumers to endorse the degree to which they feel that the brand in question is characterized by each presented attribute, and then we utilize the conditional independence properties of the recovered data to impute the existence of an underlying link structure among those attributes. This clearly differs strongly from the approach pioneered by Roedder John et al. (1996). As mentioned several times throughout the thesis, we feel that this difference can perhaps be regarded as a strength of the technique, since it derives the brand concept structure from an easily retrievable form of data that does not place undue burdens on the respondents. However, it is an indirect method, while the technique of Roedder John et al (ibid.) utilizes a direct elicitation methodology. In this sense, the techniques developed here essentially go down a somewhat different analytical pathway than does the BCM approach. Perhaps there is ground for a common structural discovery protocol in which the direct elicitation methodology pioneered by Roedder John et al. can be utilized along with the Bayesian network-based technique pursued here in order to determine a brand concept structure through a ‘triangulation’ between these differing approaches.

In addition to these potential limitations of the techniques pursued in this thesis there are also several unanswered questions that could represent fertile ground for future exploration. For example, we found (through a Markov equivalence class representation of the set of cognitive variants that can coexist within a given set of brand concept data) that there are typically cognitive subgroups of consumers, each of which may view specific links as having a direction that is opposite to that seen by other cognitive subgroups within that data set. This raises the question of how prevalent each such subgroup is, as well as the intriguing possibility of using Bayesian network-based measures to quantify the cognitive 'distance' between such subgroups. This might also allow researchers to uncover specific relations among such subgroups which relate to or expand upon other findings that are known within the sociocognitive influences literature.

An additional hypothesis is related to the work presented here concerns information propagation within a brand concept network. As we saw earlier, the sepsets (i.e., the minimal separating sets) within a clique tree representation of a Bayesian network provide a minimal "alphabet" of locations upon which one needs to condition in order to ascertain the degree and direction of information flow through the network as a whole. Therefore, this property of network structures leads to the natural hypothesis that individuals may preferentially *choose* to condition on such minimal separators (or 'sepsets') when selecting which variables to observe or manipulate (either actively or counterfactually) in order to understand the effect that one variable of interest has upon another within their targeted consumption domain, and hence these may be the variables that consumers naturally tend to focus on when constructing their consideration sets or determining what to purchase. Of course a related hypothesis would be that marketers may find it more effective to focus their advertising and communication strategies upon these variables, since these particular sets of variables serve as "linchpins" which regulate and restrict the flow of information through the network as a whole

Lastly, a potentially fascinating area for future exploration is the relation between directionally protected links and reversible links with regards to consumers' perceptions of the overall brand concept as well as the more general consumption domain in which that brand concept resides. For instance, it stands to reason that since all permissible cognitive subgroups within a data set share the same set of directionally conserved or protected links, that such protected links are apt to be more difficult to influence through managerial action. This raises the question of whether one can develop an index of relational pliability or manipulability which can quantify the relative level of influence that management may have on that particular relationship.

Furthermore, we saw that the central reversible link within the Taco Prima brand concept map was also the link which had the strongest link strength measure. It is unclear if this is coincidental, or whether there are either cognitive or strategic reasons why this should occur. In fact, as we discussed earlier, it is possible that rather than merely *indicating* where important variables are located, an inter-variable association may actually have a high link strength *because* it is a location which drives cognitive differentiation among various causal subgroups of consumers. If this is so, then such an analysis would represent an invaluable tool for managers seeking to determine the optimal brand associations and inter-association relations upon which to intervene in order to pursue various brand-related strategic objectives.

References

Aaker, D.A. (1991). *Managing Brand Equity*. *Capitalizing on the Value of a Brand Name*. New York, NY: The Free Press.

Aaker, D.A. (1996). Measuring Brand Equity Across products and Markets. *California Management Review,* 38 (Spring), 102-20.

Alba, J.W., and Hutchinson, J. W. (1987). Dimensions of Consumer Expertise. *Journal of Consumer Research,* 13, 411-454.

Allan, L.G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, 15(3), 147-149.

Anderson, J.R., and Bower, G.H. (1973). *Human Associative Memory*. Washington D.C. : V. H. Winston & Sons.

Anderson, J.R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology,* 6, 451-474.

Anderson, J.R. (1983a). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Anderson, J.R. (1983b). A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior,* 22, 261-295.

Anderson, J.R., and Reder, L.M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General,* 128, 186-197.

Babbie, E.W. (1998). *The Practice of Social Research, 8th ed.* Belmont, CA: Wadsworth.

Barnard, M. (1996). *Fashion as Communication*. New York: Routledge.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. New York: Cambridge University Press.

Barsalou, L.W. (1983). Ad hoc categories. *Memory and Cognition*, 11, 211-227.

Barsalou, L.W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 11, 629-649.

Boatwright, P. and Nunes, J.C. (2001). Reducing Assortment: An Attribute-Based Approach. *Journal of Marketing*, 65 (July), 50-63.

Boerlage, B. (1992). *Link Strengths in Bayesian Networks*. Master's thesis, Department of Computer Science, University of British Columbia.

Broniarczyk, S.M., Hoyer, W.D., and McAlister, L. (1998). Consumers' Perceptions of the Assortment Offered in a Grocery Category: The Impact of Item Reduction. *Journal of Marketing Research*, 35 (May), 166-176.

Broniarczyk, S.M. (2006). Product Assortment. In: C. Haugtvedt, P. Herr, and F. Kardes (Eds.), *Handbook of Consumer Psychology* (pp. 755-780). New York: Psychology Press.

Brownstein, S., Sirsi, A., Ward, J.C., and Reingen, P.H. (2000). Lattice Analysis in the Study of Motivation. In: Ratneshwar, S., Mick, D.G., and Huffman, C. (eds.): *The Why of Consumption*, pp. 282-302. New York: Routledge.

Buntine, W. (1991). Theory refinement on Bayesian networks. In: *Proceedings of the Seventh European Conference on Uncertainty in Artificial Intelligence*, pp. 52-60. Los Angeles: Morgan Kaufmann.

Chickering, D. (1995). A Transformational Characterization of Equivalent Bayesian Network Structures. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 87-98.

Chickering, D. (1996). Learning equivalence classes of Bayesian network structures. In: *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pp. 150-157.

Chickering, D., Geiger, D., and Heckerman, D. (1995). Learning Bayesian Networks: Search Methods and Experimental Results. In: *Proceedings of the Fifth Conference on Artificial Intelligence and Statistics*, pp. 112-128. Fort Lauderdale, FL: Society for Artificial Intelligence in Statistics.

Collins, A.M., and Loftus, E.F. (1975). A Spreading Activation Theory of Semantic Processing. *Psychological Review* 82: 407-428.

Collins, A.M., and Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior* 8: 240-248.

Dillon, W.R., Madden, T.J., and Firtle, N.H. (1987). *Marketing Research in a Marketing Environment*. Boston: Richard D. Irwin.

Druzdzel, M.J., and Simon, H.A. (1993). Causality in Bayesian Belief Networks. In: *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI-93)*, pp. 317-325. San Fransisco, CA: Morgan Kaufmann Publishers.

Druzdzel, M. (1999). SMILE: Structural modeling, inference, and learning engine and GeNie: A development environment for graphical decision-theoretic models. In: *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pp. 902-903. Palo Alto, CA: AAAI Press.

Dunning, David, and Parpal, Mary (1989). Mental addition versus subtraction in counterfactual reasoning: On assessing the impact of personal actions and life events. *Journal of Personality and Social Psychology*, 57(1), 5-15.

Ebert-Uphoff, I. (2007). *Measuring Connection Strengths and Link Strengths in Discrete Bayesian Networks*. Technical Report GT-IIC-07-01, Georgia Institute of Technology.

Eisenstein, Eric M., and Hutchinson, J. Wesley (2006). Action-Based learning: Goals and Attention in the Acquisition of Market Knowledge. *Journal of Marketing Research*, 43(2), 244-258.

Farquhar, P.H., and Herr, P.M. (1993). "The Dual Structure of Brand Associations," in *Brand Equity and Advertising: Advertising's Role in Building Strong Brands*, David A. Aaker and Alexander L. Biel, (eds.). Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 263-277.

Folkes, V.S. (1988). Recent Attribution Research in Consumer Behavior: A Review and New Directions. *Journal of Consumer Research,*14, 548-565.

Food Marketing Institute (2013). "Supermarket Facts. Industry Overview 2011 - 2012." Retrieved August 12, 2013, from http://www.fmi.org/research-resources/supermarket-facts

Geiger, D., and Pearl, J. (1990). On the Logic of Causal Models. In: Schachter, R.D., Levitt, T.S., Kanal, L.N., and Lemmer, J.F. (Eds.), *Uncertainty in Artificial Intelligence, Proceedings of the Fourth Conference*, (pp. 136-147). Amsterdam: North Holland.

Geiger, D., Verma, T., and Pearl, J. (1990). Identifying independence in Bayesian networks. *Networks*, 20 (5), 507-534.

Gilovich, T., Vallone, R., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17, 295-314.

Glazer, R., and Nakamoto, K. (1991). Cognitive Geometry: An Analysis of Structure Underlying Representations of Similarity. *Marketing Science,* 10(3), 205 -228.

Gomez-Ariza, C.J., and Bajo, M.T. (2003). Inference and integration: The fan effect in children and adults. *Memory*, 11 (6), 505-523.

Gutman, J. (1997). Means-end chains as goal hierarchies. *Psychology and Marketing* 14: 545-560.

Gwin, C.F., and Gwin, C.R. (2003). Product Attributes Model: A Tool for Evaluating Brand Positioning. *Journal of Marketing Theory and Practice*, 11(2), 30-42.

Hauser, J.R., and Simmie, P. (1981). Profit Maximizing Perceptual Positions: An Integrated Theory for the Selection of Product Features and Price. *Management Science*, 27 (1), 33-56.

Heckerman, D. (1995). A Tutorial on Learning with Bayesian Networks. Technical Report MSR-TR-95-06. Redmond, WA: Microsoft Research.

Heckerman, D., Geiger, D., and Chickering, D.M. (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*, 20, 197-243.

Henderson G.R., Iacobucci, D., and Calder, B.J. (1998). Brand diagnostics: Mapping branding effects using consumer associative networks. *European Journal of Operational Research*, 111, 306-327.

Herr, P.M., Farquhar, P.H., and Fazio, R.H. (1996). Impact of Dominance and Relatedness on Brand Extensions. *Journal of Consumer Psychology*, 5(2), 135-159.

Hoch, Stephen J., and Ha, Young-Won (1986). Consumer Learning: Advertising and the Ambiguity of Product Experience. *Journal of Consumer Research*, 13(2), 221-233.

Hoch, Stephen J., and Deighton, John (1989). Managing What Consumers Learn from Experience. *Journal of Marketing*, 53(2), 1-20.

Holbrook, Morris B., and Hirschman, Elizabeth C. (1982). The Experiential Aspects of Consumption: Consumer Fantasies, Feelings, and Fun. *Journal of Consumer Research,* 9(2), 132-140.

Holden, S.J., and Lutz, R.J. (1992). Ask Not What the Brand Can Evoke; Ask What Can Evoke the Brand? *Advances in Consumer Research*, 19, 101-107.

Holman, R.H. (1980). Clothing as Commumication: An Empirical Investigation. *Advances in Consumer Research*, 7(1), 372-377.

Huang, C., and Darwiche, A. (1996). Inference in Belief Networks: A Procedural Guide. *International Journal of Approximate Reasoning*, 15, 225-263.

Huber, J., and Holbrook, M.B. (1979). Using Attribute Ratings for Product Positioning: Some Distinctions Among Compositional Approaches. *Journal of Marketing Research*, 16 (November), 507-515.

Hunt, Shelby D. (1991). *Modern Marketing Theory: Critical Issues in the Philosophy of Marketing Science*. Cincinnati, OH: South-Western Publishing Company.

Hutchinson, J. Wesley, and Eisenstein, Eric M. (2008). Consumer Learning and Expertise. In Haugtvedt, C.; Herr, P.; and Kardes, F. (Eds.), *Handbook of Consumer Psychology* (pp. 103-132). New York: Taylor and Francis.

Jitnah, N. (1999). *Using Mutual Information for Approximate Evaluation of Bayesian Networks*. Doctoral Thesis, School of Computer Science, Monash University, Clayton, Victoria, Australia.

Kahneman, D., and Tversky, A. (1982). The simulation heuristic. In D.Kahneman, P.Slovic, & A.Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.

Kassarjian, H.H. (1974). Projective Methods, in Robert Ferber (ed.), *Handbook of Marketing Research*, pp. 85-100. New York: McGraw-Hill.

Keller, K. (1993). Conceptualizing, Measuring, and Managing Customer-Based Brand Equity. *Journal of Marketing*, 57(1), 1-22.

Kelley, H. (1973). The processes of causal attribution. *American Psychologist, 28*(2), 107-128.

Kim, J. and Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference systems. In: *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, pp. 190-193. San Francisco: Morgan Kaufmann.

Kjaerulff, U., and Madsen, A. (2008). *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. New York: Springer Verlag.

Klahr, D. (1970). A Study of Consumer's Cognitive Structure for Cigarette Brands. *Journal of Business*, 43 (April), 190-204.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211–228.

Koller, D., and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: Massachusetts Institute of Technology Press.

Koski, T., and Noble, J. (2009). *Bayesian Networks: An Introduction*. New York: Wiley.

Krishnamurthy, P., and Sivaraman, A. (2002). Counterfactual Thinking and Advertising Responses. *Journal of Consumer Research*, 28(4), 650-658.

Lagnado, D. A., and Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 856-876.

Lagnado, D.A., Waldmann, M. R., Hagmayer, Y., & Sloman, S.A. (2007). Beyond Covariation: Cues to Causal Structure. In Gopnik, A., & Schultz, L. (Eds.), *Causal Learning: Psychology, Philosophy, and Computation.* New York: Oxford University Press.

Lakoff, George, and Johnson, Mark (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lauritzen, S. (1999). Causal inference from graphical models. In: D. Barnsdorf-Nielsen and C. Klupenberg, eds., *Complex Stochastic Systems*, pp.141-165 Baton Rouge, LA: Chapman and Hall.

Lauritzen, S., Dawid, A., Larsen, B., and Leimer, H. (1990). Independence properties of directed Markov fields. *Networks*, 20(5), 491-505.

Lei, J., Dawar, N., and Lemmink, J. (2008). Negative Spillover in Brand Portfolios: Exploring the Antecedents of Asymmetric Effects. *Journal of Marketing*, 72 (May), 111-123.

Levy, S.J. (1985). Dreams, fairy tales, animals, and cars. *Psychology and Marketing*, 2(2): 67-81.

Levy, S.J. (2006). History of qualitative research methods in marketing. In: Russel W. Belk (ed.), *Handbook of Qualitative Research Methods in Marketing*. Northampton, MA: Edward Elgar Publishing, pp. 3-16.

Lewis, C.H., and Anderson, J.R. (1976). Interference with real world knowledge. *Cognitive Psychology*, 8, 311-335.

MacKay, D.J.C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.

Markman, K. D., Lindberg, M. J., Kray, L. J., and Galinsky, A. D. (2007). Implications of Counterfactual Structure for Creative Generation and Analytical Problem Solving. *Personality and Social Psychology Bulletin*, 33(3), 312-324.

McCracken, G., and Roth, V. (1989). Does clothing have a code? Empirical findings and theoretical implications in the study of clothing as a means of communication. *International Journal of Research in Marketing*, 6, 13-33.

McGill, Ann L. (2000). Counterfactual reasoning in causal judgments: Implications for marketing. *Psychology and Marketing* 17(4): 323-343.

McGuire, M.J., and Maki, R. (2001). When Knowing More Means Less: The Effect of Fan on Metamemory Judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(5): 1172-1179.

Meek, C. (1995). Causal inference and causal explanation with background knowledge. In P. Besnard and S. Hank (Eds.), *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence*. San Fransisco, CA: Morgan Kaufman, pp. 403-410.

Morgan, N. and Purnell, J. (1969). Isolating Openings for New Products in a Multidimensional Space. *Journal of the Marketing Research Society*, 11 (July), 245-266.

Meyers-Levy, J., and Maheswaran, D. (1991). Exploring differences in males' and females' processing strategies. *Journal of Consumer Research*, 18, 63-70.

Meyers-Levy, J., Sternthal, B. (1991). Gender difference in the use of message cues and judgements. *Journal of Marketing Research*, 28, 84-96.

Nedungadi, P. (1990). Recall and Consumer Consideration Sets: Influencing Choice without Altering Brand Evaluations. *Journal of Consumer Research*, 17 (December), 263-276.

Nicholson, A., and Jitnah, N. (1998). Using Mutual Information to Determine Relevance in Bayesian Networks. In *5th Pacific Rim International Conference on Artificial Intelligence (PRICAI '98)*, pages 399-410. Singapore: Springer Verlag.

Park, C. W., Jaworski, B.J., and MacInnis, D.J. (1986). Strategic Brand Concept-Image Management. *Journal of Marketing*, 50 (October), 135-45.

Park, C.W., Milberg, S., and Lawson, R. (1991). Evaluation of Brand Extensions: The Role of Product-Level Similarity and Brand Concept Consistency. *Journal of Consumer Research*, 18 (September), 185-93.

Pearl, Judea (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann Publishers.

Pearl, Judea (2000). *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Pearl, J., Geiger, D., and Verma, T. (1989). Conditional Independence and its Representations. *Kybernetika*, 25(Supplement), 33-44

Pennington, N., and Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition,* 49, 123-163.

Radvansky, G.A. (1999). The fan effect: A tale of two theories. *Journal of Experimental Psychology: General*, 2, 198-206.

Reynolds, T.J., and Gutman, J. (1988). Laddering theory: method, analysis, and interpretation. *Journal of Advertising Research*, 28, 11-31.

Roedder John, D., Loken, B., Kim, K., and Basu Monga, A. (2006). Brand Concept Maps: A Methodology for Identifying Brand Association Networks. *Journal of Marketing Research,* 43(4), 549-563.

Roese, Neal J., and Olson, James M. (1996). Counterfactuals, Causal Attributions, and the Hindsight Bias: A Conceptual Integration. *Journal of Experimental Social Psycholog*y, 32, 197-227.

Rook, D.W. (2006). Let's pretend: projective methods reconsidered. In: Russel W. Belk (ed.), *Handbook of Qualitative Research Methods in Marketing*, pp. 143-155. Northampton, MA: Edward Elgar Publishing.

Ruiz-Primo, M.A., and Shavelson, R.J. (1996). Problems and Issues in the Use of Concept Maps in Science Assessment. *Journal of Research in Science Teaching*, 33 (August), 569-600.

Shannon, C.E., and Weaver, W. (1949). *The Mathematical Theory of Communication*. Champaign-Urbana: University of Illinois Press.

Shocker, A.D., and Srinivasan, V. (1974). A Consumer-Based methodology for the Identification of New Product Ideas. *Management Science*, 20(6), 921-937.

Shocker, A.D., and Srinivasan, V. (1979). Multiattribute Approaches for Product Concept Evaluation and Generation: A Critical Review. *Journal of Marketing Research,* 16(2), 159-180.

Sirsi, Ajay K., James C. Ward, and Peter H. Reingen (1996), Microcultural Analysis of Variation in Sharing of Causal Reasoning About Behavior, *Journal of Consumer Research*, 22 (4), 345-372

Sloman, Steven A. (2005). *Causal models: How people think about the world and its alternatives*. New York: Oxford University Press.

Spirtes, Peter; Glymour, Clark; and Scheines, Richard (2000). *Causation, Prediction, and Search*. Cambridge, MA: Massachusetts Institute of Technology Press.

Steyvers, M., Tenenbaum, J., Wagenmakers, E., and Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, 27, 453-489.

Summers, J., and MacKay, D. (1976). On the Validity and Reliability of Direct Similarity Judgments. *Journal of Marketing Research*, 13, 289-295.

Summerville, A., and Roese, N. J. (2008). Dare to compare: Fact-based versus simulation-based comparison in daily life. *Journal of Experimental Social Psychology*, 44, 664–671.

Tversky, A. (1969). Intransitivity of Preferences. *Psychological Review,* 76(1), 31-48.

Tversky, A. (1977). Features of Similarity. *Psychological Review*, 84(4), 327-52.

Tversky, A., and Gati, I. (1978). Studies of Similarity, in *Cognition and Categorization*, E. Rosch and B.B. Lloyd, eds. Hillsdale, NJ: Lawrence Erlbaum Associates, p. 79-98.

Tversky, A., and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131.

Tversky, A., and Kahneman, D. (1980). Causal schemata in judgments under uncertainty. In: M. Fishbein (Ed.), *Progress in Social Psychology*, pp. 49-72. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tversky, A., and Kahneman, D. (1982). Causal schemas in judgments under uncertainty. In D. Kahnemann, P. Slovic, and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 117-128). Cambridge: Cambridge University Press.

Ulhaque, E., and Bahn, K.D. (1992). A Spreading Activation Model of Consumers' Asymmetric Similarity Judgment. *Advances in Consumer Research,* 19, 782-786.

Verma, T., and Pearl, J. (1988). Causal Networks: Semantics and Expressiveness. In *Proceedings of the Fourth Conference on Uncertainty in Artificial Intelligence*, pp. 352-359. San Francisco: Morgan Kaufmann.

Verma, T., and Pearl, J. (1990). Equivalence and Synthesis of Causal Models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence,* pp. 220-227. San Francisco: Morgan Kaufmann.

Waldmann, M.R., and Holyoak, K.J. (1990). Can causal induction be reduced to associative learning ? In: Proceedings of the 12th annual conference of the cognitive science society (pp. 190-197). Hillsdale, NJ: Erlbaum.

Weiner, B. (2000). Attributional Thoughts about Consumer Behavior. *Journal of Consumer Research*, 27(3), 382-387.

Wiener, J.L., and Mowen, J.C. (1986). Source Credibility: On the Independent Effects of Trust and Expertise. In R. Lutz (ed.), *Advances in Consumer Research*, 13: 306-310.

Wellman, Michael, and Henrion, Max (1993). Explaining "Explaining Away". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3), 287-292.

Wells, G.L. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology* 62(5): 739-752.

Wells, W.D. (1975). Psychographics: A Critical Review. *Journal of Marketing Research* 12 (May): 196-213.

Wolford, G. (1971). Function of distinct associations for paired-associate performance. *Psychological Review*, 78, 303-313.

Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation.* New York: Oxford University Press.

Woodward, James (2007). Interventionist Theories of Causation in Psychological Perspective. In: A. Gopnik & L. Schulz (Eds.), *Causal Learning: Psychology, Philosophy, and Computation*. New York: Oxford University Press.

Zaltman, G., and Coulter, R. (1995). Seeing the voice of the customer: metaphor-based advertising research. *Journal of Advertising Research*, 35(4), 35-51.

Zaltman, G. (1997). Rethinking marketing research: putting people back in. *Journal of Marketing Research*, 34, 424-37.

APPENDIX A

D-SEPARATION IN CAUSAL AND DIAGNOSTIC CHAINS

A chain connection (e.g.,) consists of three random variables *X*, *Y*, and Z for which *X* directly influences *Y* and for which *Y* directly influences *Z*, but for which there is no *direct* influence between *X* and *Z*. Given this graphical arrangement, the joint probability of the three random variables *X*, *Y*, and *Z* is given by:

$$p_{X,Y,Z} = p_{Z\,|\,Y} \cdot p_{Y\,|\,X} \cdot p_X \tag{35}$$

We will discuss this form of connection with regards to multinomial data only. Further details can be found in Koski and Noble (2009) and Koller and Friedman (2009).

As stated in the thesis, we assert that conditioning on the central variable *Y* of the chain connection $\mathrm{X} \rightarrow \mathrm{Y} \rightarrow \mathrm{Z}$ blocks communication between the terminal variables *X* and Z within that connection. To prove this assertion, note that by the basic probability calculus, *any* triple (*X*, *Y*, *Z*) of random variables will have a joint probability distribution that factors according to the form $p_{X,Y,Z} = p_{X,Z\,|\,Y} \cdot p_Y$ (i.e., the joint probability factors as the product of the conditional and marginal distributions). Thus we can isolate the joint probability of the terminal vertices *X* and *Z* in this connection, as conditioned on the central variable *Y*, as follows:

$$P_{X,Z\,|\,Y} = \frac{P_{X,Y,Z}}{P_Y} \tag{36}$$

The above expression is valid for any triple of random variables for which $p_Y \neq 0$. However, we can also factor the joint probability of X, Y, and Z in the numerator of this expression according to the specific topology of the chain connection, yielding the following expression specific to this particular connection topology:

$$p_{X,Z\,|\,Y} \;=\; \frac{p_{\,Z\,|\,Y} \cdot p_{\,Y\,|\,X} \cdot p_{\,X}}{p_Y} \tag{37}$$

Assuming $p_X \neq 0$, we can now utilize Bayes' theorem to re-express the central quantity $p_{\,Y|X}$ in the numerator of the above expression, thereby yielding the transformed expression shown below:

$$p_{X,Z\,|\,Y} \;=\; \frac{p_{\,Z\,|\,Y} \cdot \left[\frac{p_{X|Y} \cdot p_Y}{p_X}\right] \cdot p_{\,X}}{p_Y} \tag{38}$$

Cancelling common factors of p_Y and common factors of p_X yields the reduced expression:

$$p_{X,Z\,|\,Y} \;=\; p_{\,Z\,|\,Y} \cdot \; p_{\,X\,|\,Y} \tag{39}$$

and hence the terminal variables X and Z are independent in this topology when conditioned on the central variable Y of the topology.

Alternatively, one can also demonstrate this independence of X and Z conditional on Y by examining the distribution of Z conditional on X and Y as follows:

$$p_{\,Z\,|\,X,Y} \;=\; \frac{p_{X,Y,Z}}{p_{X,Y}} \;=\; \frac{p_{\,Z|Y} \cdot p_{\,Y|X} \cdot p_{\,X}}{p_{\,Y|X} \cdot p_{\,X}} \;=\; p_{\,Z|Y} \tag{40}$$

Of course, here we once again used the specific factorization of the joint distribution of X, Y, and Z along the chain topology when decomposing the numerator of this expression. Therefore, we obtain another justification of the fact that communication between the terminal variables X and Z of the chain topology is blocked (i.e., variables X and Z are rendered independent) when conditioned on the central variable Y of the topology.

Forward propagation of information along the chain connection :

In addition to the above analysis showing the induced conditional independence of the terminal variables X and Z in the chain topology, we can also utilize this factorization of the joint distribution in order to ***examine how messages are passed along a chain connection***. Specifically, by marginalizing the factored distribution over Y and utilizing the factorization specific to this connection topology, we obtain:

$$\begin{aligned} p_{X,Z}(X,Z) &= \sum_{y} p_{Z|Y}(Z|y)\, p_{Y|X}(y|X)\, p_X(X) \\ &= p_X(X) \cdot \sum_{y} p_{Z|Y}(Z|y)\, p_{Y|X}(y|X) \end{aligned} \tag{41}$$

Essentially, the summation $\sum_y \left[p_{Z|Y}(Z|y)\, p_{Y|X}(y|X) \right]$ in the last factor is playing the role of $p_{Z|X}(Z|X)$ in the expression $p_{X,Z}(X,Z) = p_X(X) \cdot p_{Z|X}(Z|X)$. Since $p_{Z|X}(Z|X) = \sum_y p_{Z|Y}(Z|y)\, p_{Y|X}(y|X) \neq p_Z(Z)$, we have that $p_{X,Z}(X,Z) \neq p_X(X) \cdot p_Z(Z)$ in this topology, and X and Z are therefore dependent in this connection structure. *Furthermore, the expression* $p_{Z|X}(Z|X) = \sum_y p_{Z|Y}(Z|y)\, p_{Y|X}(y|X)$ *shows the explicit form of the dependence of* Z *upon* X *within the chain connection.*

We can also derive the result of conditioning the joint distribution of X, Y, and Z on the value of Y by marginalizing the specific form of the factorization of the joint distribution of X, Y, and Z along the chain connection as follows:

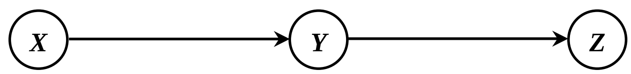
$$
\begin{aligned}
p_{X,Z|Y}(X,Z|y) &= \frac{p_{Z|Y}(Z|y)\, p_{Y|X}(y|X)\, p_X(X)}{\sum_x \sum_z\, p_{Z|Y}(z|y)\, p_{Y|X}(y|x)\, p_X(x)} \\
&= \frac{p_{Z|Y}(Z|y)\, p_{Y|X}(y|X)\, p_X(X)}{\sum_x p_{Y|X}(y|x)\, p_X(x) \sum_z\, p_{Z|Y}(z|y)} \\
&= \frac{p_{Y|X}(y|X)\, p_X(X)}{\sum_x p_{Y|X}(y|x)\, p_X(x)\, \cdot\, 1} \cdot p_{Z|Y}(Z|y) \\
&= \frac{p_{Y,X}(y,X)}{p_Y(y)} \cdot p_{Z|Y}(Z|y) \\
&= p_{X|Y}(X|y) \cdot p_{Z|Y}(Z|y)
\end{aligned}
\tag{42}
$$

This computation shows that instantiation of the central variable Y within the chain connection renders variables X and Z conditionally independent.

However, it is much more instructive to more closely trace the detailed mechanism by which X influences Z (and conversely, by which Z influences X) within a chain connection, and to observe the manner in which these paths of influence are blocked once the central variable Y within this chain connection becomes instantiated.

Such an analysis also facilitates the study of message-passing schemes within the Bayesian network.

As long as variable Y remains *uninstantiated*, variables X and Z are able to "communicate" with each other. For example, suppose that the variables X, Y, and Z are each binary, with probabilities parameterized as shown in Figure 33.



p_X	x_1	x_2
	λ	$1-\lambda$

$p_{Y\|X}$	y_1	y_2
x_1	α	$1-\alpha$
x_2	β	$1-\beta$

$p_{Z\|Y}$	z_1	z_2
y_1	δ	$1-\delta$
y_2	γ	$1-\gamma$

Figure 33. Distribution of binary variables in the chain connection X → Y → Z

Given no additional information, the probability that variable Z will take a specific value (say, $Z = z_1$) in this topology can be computed as follows :

$$p_Z(z_1) \;=\; p_{Z|Y}(z_1|y_1) \cdot p_Y(y_1) \;+\; p_{Z|Y}(z_1|y_2) \cdot p_Y(y_2)$$

$$= \; p_{Z|Y}(z_1|y_1)\big[\, p_{Y|X}(y_1|x_1) \cdot p_X(x_1) + \; p_{Y|X}(y_1|x_2) \cdot p_X(x_2)\big]$$

$$+\, p_{Z|Y}(z_1|y_2)\big[\, p_{Y|X}(y_2|x_1) \cdot p_X(x_1) + \; p_{Y|X}(y_2|x_2) \cdot p_X(x_2)\big] \tag{43}$$

After some algebraic simplification, this expression gives:

$$p_Z(z_1) = \lambda(\alpha - \beta)(\delta - \gamma) + \beta(\delta - \gamma) + \gamma \tag{44}$$

Similarly, we can express $p_Z(z_2)$ in terms of the probabilities given in this chain connection topology, and after simplification we of course find that:

$$p_Z(z_2) = 1 - p_Z(z_1) = 1 - \gamma - \beta(\delta - \gamma) - \lambda(\alpha - \beta)(\delta - \gamma) \tag{45}$$

Since these expressions depend on λ, we see that the distribution of the terminal variable in the chain topology is sensitive to the probability distribution of the initial variable in the chain.

However, we can say even more. Suppose that we discover that the variable X has become instantiated as, say, $X = x_1$. Then this 'signal' will become propagated through the chain topology in the sense that we can now condition the distribution of Z on this instantiated value of X. Specifically, using the given probabilities, we can compute this conditional probability as:

$$\begin{aligned} p_{Z|X}(z_1|x_1) &= p_{Z|YX}(z_1|\, y_1, x_1) \cdot p_{Y|X}(y_1|x_1) + p_{Z|YX}(z_1|\, y_2, x_1) \cdot p_{Y|X}(y_2|x_1) \\ &= p_{Z|Y}(z_1|y_1) \cdot p_{Y|X}(y_1|x_1) + p_{Z|Y}(z_1|y_2) \cdot p_{Y|X}(y_2|x_1) \end{aligned} \tag{46}$$

and after some algebraic simplification, this expression reduces to the simpler form $p_{Z|X}(z_1|x_1) = \alpha(\delta - \gamma) + \gamma$. Similarly, $p_{Z|X}(z_2|x_1)$ can be expressed as $1 - \gamma - \alpha(\delta - \gamma)$. Obviously, the information that $X = x_1$ has drastically affected the distribution of the variable Z in a process that we can term *forward propagation* through the chain connection.

Interestingly, the only way that the posterior distribution of variable Z (as conditioned on the information that $X = x_1$) can be equal to the prior distribution of Z would be for the quantity $\alpha(\delta - \gamma) + \gamma$ to equal the quantity $\lambda(\alpha - \beta)(\delta - \gamma) + \beta(\delta - \gamma) + \gamma$. Equating these two expressions and reducing yields :

$$\begin{aligned} \alpha(\delta - \gamma) + \gamma &= \lambda(\alpha - \beta)(\delta - \gamma) + \beta(\delta - \gamma) + \gamma \\ \Leftrightarrow \quad 0 &= \lambda(\alpha - \beta)(\delta - \gamma) + \beta(\delta - \gamma) - \alpha(\delta - \gamma) \\ \Leftrightarrow \quad 0 &= \lambda(\alpha - \beta)(\delta - \gamma) - (\alpha - \beta)(\delta - \gamma) \\ \Leftrightarrow \quad 0 &= (\lambda - 1)(\alpha - \beta)(\delta - \gamma) \end{aligned} \tag{47}$$

Hence the only way for the posterior distribution of Z conditioned on $X = x_1$ to equal the prior distribution would be for: (i) α to equal β (in which case $p_{Y|X}(y_1|x_1) = p_{Y|X}(y_1|x_2)$, meaning that variables X and Y would be independent), (ii) for δ to equal γ (in which case $p_{Z|Y}(z_1|y_1) = p_{Z|Y}(z_1|y_2)$, meaning that variables Y and Z would be independent), or (iii) for λ to equal 1 (in which case $p_X(x_1) = 1$). Of course if either X and Y, or Y and Z are independent, then the respective link in the chain topology diagram would be absent, meaning that the diagram (and hence the factorization along the chain topology) would be invalid (i.e., an "unfaithful graphical representation"). On the other hand, if $p_X(x_1) = 1$ then the "information" that X has been instantiated as x_1 would be no new information at all, and we would be conditioning on information that is already present in the diagram. *Therefore, as long as the central variable Y in the chain topology is uninstantiated, all changes to the distribution of the initial variable X* <u>*must*</u> <u>*be*</u> *propagated forward through Y and affect the distribution of the terminal variable Z.*

Now if the central variable Y in the chain topology is *instantiated* to a specific value, then this process of forward propagation through the chain connection will be ineffective. One can see this since in the presence of an instantiation such as $Y = y_1$, the structure of the chain topology will naturally reduce both $p_{Z|X}(z_1|x_1)$ and $p_Z(z_1)$ to the quantity $p_{Z|Y}(z_1|y_1)$. Alternatively, one could incorporate the information that $Y = y_1$ by letting $\alpha = \beta = 1$ in the previously derived factorization along the chain structure, thereby reducing the expression $p_{Z|X}(z_1|x_1) = \alpha(\delta - \gamma) + \gamma$ as well as the expression $p_Z(z_1) = \lambda(\alpha - \beta)(\delta - \gamma) + \beta(\delta - \gamma) + \gamma$, to merely γ (which corresponds to $p_{Z|Y}(z_1|y_1)$ in the above diagram).

<u>Reverse propagation of information along the chain connection</u> :

The above discussion shows that the instantiation of the variable Y in the chain connection $X \to Y \to Z$ blocks the ability of a signal at X from altering the distribution of Z. Now consider the *reverse* process: a signal arriving at variable Z in this chain topology is propagated through the chain all the way back to X, this time in a process that we term *reverse* (or '*backward*') *propagation*. To see how this influence is manifested, assume that variable Z becomes instantiated as, say, $Z = z_1$. We can then use Bayes' theorem to propagate this change to the variable Y, giving:

$$p_{Y|Z}(y_1|z_1) \; = \; \frac{p_{Z|Y}(z_1|y_1) \cdot p_Y(y_1)}{p_Z(z_1)} \tag{48}$$

Following this, we can propagate the change one level further to the variable X by once again employing Bayes' theorem:

$$p_{X|Z}(x_1|z_1) \; = \; \frac{p_{Z|X}(z_1|x_1) \cdot p_X(x_1)}{p_Z(z_1)} \tag{49}$$

Note that in order to obtain the values of $p_{Z|X}(z_1|x_1)$ and $p_Z(z_1)$ within this reverse propagation expression, we have to use the *forward propagation* step described above. Employing these techniques in tandem (and once again using the parameters given in the above diagram) yields the expression:

$$\begin{aligned} p_{X|Z}(x_1|z_1) \;&=\; \frac{p_{Z|X}(z_1|x_1) \cdot p_X(x_1)}{p_Z(z_1)} \\ &=\; \frac{[\,\alpha(\delta-\gamma) \;+\; \gamma\,] \cdot \lambda}{\lambda(\alpha-\beta)(\delta-\gamma) \;+\; \beta(\delta-\gamma) + \;\gamma} \end{aligned} \tag{50}$$

Therefore the information that $Z = z_1$ has altered the probability $p_X(x_1)$ from its prior value of λ to the posterior value given here.

In fact, the only way for the information that $Z = z_1$ to *not* alter the distribution of *X* is for $p_{X|Z}(x_1|z_1)$ to equal $p_X(x_1)$, and this will only occur under the condition that:

$$\lambda = \frac{[\,\alpha(\delta - \gamma) + \gamma\,] \cdot \lambda}{\lambda(\alpha - \beta)(\delta - \gamma) + \beta(\delta - \gamma) + \gamma} \tag{51}$$

Simplification of this equation results in the expression:

$$\lambda^2(\alpha - \beta)(\delta - \gamma) + (\beta\lambda - \alpha\lambda)(\delta - \gamma) = 0 \tag{52}$$

and hence that:

$$(\lambda^2 - \lambda)(\alpha - \beta)(\delta - \gamma) = 0 \tag{53}$$

Therefore the only way that the instantiation of *Z* would *not* alter the distribution of *X* via reverse propagation in this topology is for either X and Y to be independent (i.e., $\alpha = \beta$), for Y and Z to be independent (i.e., $\delta = \gamma$), or for the value of *X* to already be a certainty (i.e., $\lambda^2 = \lambda$, hence either $p_X(x_1) = 1$ or $p_X(x_1) = 0$, so either $X = x_1$ or $X = x_2$ respectively).

Once again, if we assume that the central variable *Y* in the chain topology is *instantiated* to a specific value, then this process of reverse propagation through the chain connection will be *blocked*. This is fairly obvious based on the topology since *X* can depend on *Z* only through *Y*. Alternatively, we can demonstrate this blocking effect by incorporating the assumption that $Y = y_1$ via setting the value of parameters α and β both to 1. Through this assumption (and once again using the parameters given above), we have that $p_{X|ZY}(x_1|z_1, y_1) = \frac{[\,\alpha(\delta-\gamma) + \gamma\,]\cdot\lambda}{\lambda(\alpha-\beta)(\delta-\gamma) + \beta(\delta-\gamma)+\gamma} = \frac{[\,1(\delta-\gamma) + \gamma\,]\cdot\lambda}{\lambda(1-1)(\delta-\gamma) + 1(\delta-\gamma)+\gamma} =$

$\frac{\delta\cdot\lambda}{\delta} = \lambda$ (i.e., under the assumption that $= y_1$, the conditional probability $p_{X|Z}(x_1|z_1)$ reduces to $p_X(x_1)$). Similarly, under the assumption that $= y_2$, parameters α and β both reduce to 0, so we have that $p_{X|ZY}(x_1|z_1, y_2) = \frac{[\,\alpha(\delta-\gamma) + \gamma\,]\cdot\lambda}{\lambda(\alpha-\beta)(\delta-\gamma) + \beta(\delta-\gamma)+\gamma} =$

$\frac{[\,0(\delta-\gamma) + \gamma\,]\cdot\lambda}{\lambda(0-0)(\delta-\gamma) + 0(\delta-\gamma)+\gamma} = \frac{\gamma\cdot\lambda}{\gamma} = \lambda$ (hence under the assumption that $Y = y_2$, the conditional probability $p_{X|Z}(x_1|z_1)$ once again reduces to $p_X(x_1)$).

The above analyses show that instantiation of the central variable *Y* in the chain topology $X \rightarrow Y \rightarrow Z$ does indeed block communication between the two terminal vertices *X* and *Z*, but *not in a symmetric manner*. In the 'reverse' direction, the instantiation of *Y* as either y_1 or y_2 reduces the conditional probability $p_{X|Z}(x_1|z_1)$ to $p_X(x_1)$, an effect which can only be described as classical 'blocking': given *Y*'s instantiation, the value of *Z* has absolutely no effect on *X*'s distribution. However, in the 'forward' direction, the effect of instantiating the central variable *Y* does not so much block the effect of *X* upon *Y*, but rather 'overrides' that influence: the instantiation of *Y* as y_1 reduces both $p_{Z|X}(z_1|x_1)$ and $p_Z(z_1)$ to $p_{Z|Y}(z_1|y_1)$, and the instantiation of *Y* as y_2 reduces both $p_{Z|X}(z_1|x_1)$ and $p_Z(z_1)$ to $p_{Z|Y}(z_1|y_2)$.

In summary, for a 'chain' type connection $X \rightarrow Y \rightarrow Z$, the two terminal variables *X* and *Z* communicate as long as the intervening 'connector' variable *Y* is *uninstantiated*, and are rendered conditionally independent of each other (the communication between them is 'broken') once the intervening variable *Y* becomes instantiated.

APPENDIX B

D-SEPARATION WITHIN A COMMON CAUSE CONNECTION

A fork, or 'common cause' connection $X \leftarrow Y \rightarrow Z$ consists of three random variables X, Y, and Z for which Y directly influences both X and Z, but for which there is no *direct* influence between X and Z. Given this graphical structure, the joint probability of the three random variables X, Y, and Z is given by the expression: $p_{X,Y,Z} = p_{Z\mid Y} \cdot p_{X\mid Y} \cdot p_Y$ Once again, we will discuss this form of connection with regards to multinomial data only. Further details can be found in Koski and Noble (2009) and Koller and Friedman (2009).

First we will show that conditioning on the central variable Y blocks communication between the terminal variables X and Z. As was the case within the chain connection discussed previously, we can write that $p_{X,Y,Z} = p_{X,Z\mid Y} \cdot p_Y$ for any joint distribution of three variables X, Y, and Z, and if $p_Y \neq 0$ we can once again rearrange this factorization into the form:

$$p_{X,Z\mid Y} = \frac{p_{X,Y,Z}}{p_Y} \tag{54}$$

This expression is valid for any triple of random variables for which $p_Y \neq 0$. However, we can also factor the joint probability of X, Y, and Z in the numerator of this expression according to the specific topology of the fork connection, yielding the following expression specific to the 'fork' connection topology:

$$p_{X,Z\mid Y} = \frac{p_{Z\mid Y} \cdot p_{X\mid Y} \cdot p_Y}{p_Y} \tag{55}$$

Cancellation of the redundant p_Y factor yields the reduced factorization $p_{X,Z\,|\,Y} = p_{Z\,|\,Y} \cdot p_{X\,|\,Y}$ and therefore the terminal variables X and Z in the fork topology are conditionally independent given the value of the central variable Y of this topology.

Alternatively, one can also demonstrate this independence of X and Z conditional on Y by examining the distribution of Z conditional on X and Y as follows:

$$p_{Z\,|\,X,Y} = \frac{p_{X,Y,Z}}{p_{X,Y}} = \frac{p_{Z|Y} \cdot p_{X|Y} \cdot p_Y}{p_{X|Y} \cdot p_Y} = p_{Z|Y} \tag{56}$$

Of course, here we once again used the specific factorization of the joint distribution of X, Y, and Z along the ‘common cause’ (or ‘fork’) connection when decomposing the numerator of this expression. Therefore, we obtain another justification of the fact that communication between the terminal variables X and Z of this connection topology is blocked (i.e., variables X and Z are rendered independent) when conditioned on the central variable Y of the topology.

Furthermore, the *symmetry of the roles of variables X and Z* in the fork connection can be explicitly demonstrated by conditioning X on Y and Z (as opposed to conditioning Z on X and Y as was done above). Applying the factorization of the joint distribution along the fork connection then yields:

$$p_{X\,|\,Y,Z} = \frac{p_{X,Y,Z}}{p_{Y,Z}} = \frac{p_{Z|Y} \cdot p_{X|Y} \cdot p_Y}{p_{Z|Y} \cdot p_Y} = p_{X|Y} \tag{57}$$

Message passing along the common cause connection :

In addition to the above analysis showing the induced conditional independence of the terminal variables X and Z in the fork topology, we can also utilize this factorization of the joint distribution in order to examine how messages are passed along such a connection. Specifically, *prior to the instantiation of any variables*, the joint probability of X and Z is obtained by marginalizing the factored distribution X, Y, and Z over the various values of the central variable Y in the fork connection:

$$p_{X,Z}(X,Z) = \sum_{y} p_{Z|Y}(Z|y)\, p_{X|Y}(X|y)\, p_Y(y) \tag{58}$$

Since y is the variable of summation, we cannot remove the third factor in the summand, and hence X and Z are not independent in this connection topology. However, conditioning the joint distribution of X and Z on an instantiated value of Y will result in X and Z being (conditionally) independent in this topology, as shown below:

$$\begin{aligned}
p_{X,Z|Y}(X,Z|y) &= \frac{p_{Z|Y}(Z|y)\, p_{X|Y}(X|y)\, p_Y(y)}{\sum_x \sum_z\, p_{Z|Y}(z|y)\, p_{X|Y}(x|y)\, p_Y(y)} \\
&= \frac{p_Y(y)\, p_{Z|Y}(Z|y)\, p_{X|Y}(X|y)}{p_Y(y) \cdot \sum_x \sum_z\, p_{Z|Y}(z|y)\, p_{X|Y}(x|y)} \\
&= \frac{p_{Z|Y}(Z|y)\, p_{X|Y}(X|y)}{\sum_x p_{X|Y}(x|y)\, \sum_z p_{Z|Y}(z|y)} \\
&= p_{Z|Y}(Z|y) \cdot p_{X|Y}(X|y)
\end{aligned} \tag{59}$$

However, once again it will be much more instructive to actually trace the detailed mechanism by which X influences Z within such a connection, and to observe the manner in which these paths of influence are blocked once the central variable Y within this topology becomes instantiated.

As long as variable Y remains *uninstantiated*, variables X and Z are able to "communicate" with each other. For example, once again suppose that the variables X, Y, and Z are each binary, with probabilities parameterized as follows:

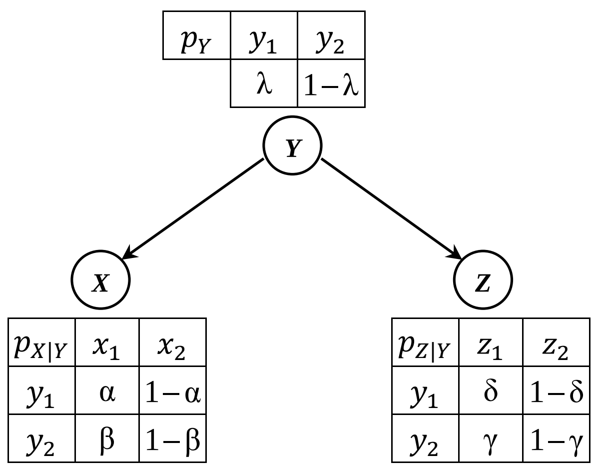


Figure 34. Distribution of binary variables in a common cause connection

To see the effect that instantiation of variable X has on variable Z, let us assume that X has the value x_1 and then compute the conditional probability $p_{Z|X}(z_1|x_1)$ within this topology. This yields the following expression:

$$p_{Z|X}(z_1|x_1) = p_{Z|YX}(z_1|\, y_1, x_1) \cdot p_{Y|X}(y_1|x_1) + p_{Z|YX}(z_1|\, y_2, x_1) \cdot p_{Y|X}(y_2|x_1) \tag{60}$$

However, in this topology variable Y once again intervenes between variables X and Z, meaning that $p_{Z|YX}(z_1|\, y_1, x_1)$ can be reduced to just $p_{Z|Y}(z_1|y_1)$, and $p_{Z|YX}(z_1|\, y_2, x_1)$ reduces to just $p_{Z|Y}(z_1|y_2)$. Making these substitutions, and using Bayes' theorem to re-express $p_{Y|X}(y_1|x_1)$ and $p_{Y|X}(y_2|x_1)$ yields:

$$p_{Z|X}(z_1|x_1) \;=\; p_{Z|Y}(z_1|y_1)\,\frac{p_{X|Y}(x_1|y_1)\,p_Y(y_1)}{p_X(x_1)} \;+\; p_{Z|Y}(z_1|y_2)\,\frac{p_{X|Y}(x_1|y_2)\,p_Y(y_2)}{p_X(x_1)} \tag{61}$$

Furthermore, since variable X receives information from variable Y, the two occurrences of $p_X(x_1)$ in this expression can be expanded using the law of total probability to reflect this dependence, giving:

$$\begin{aligned} p_{Z|X}(z_1|x_1) \;=\; & p_{Z|Y}(z_1|y_1)\,\frac{p_{X|Y}(x_1|y_1)\,p_Y(y_1)}{p_{X|Y}(x_1|y_1)\,p_Y(y_1) + p_{X|Y}(x_1|y_2)\,p_Y(y_2)} \\ & + \; p_{Z|Y}(z_1|y_2)\,\frac{p_{X|Y}(x_1|y_2)\,p_Y(y_2)}{p_{X|Y}(x_1|y_1)\,p_Y(y_1) + p_{X|Y}(x_1|y_2)\,p_Y(y_2)} \end{aligned} \tag{62}$$

Using the parameters assumed in the above diagram, this expression reduces to:

$$p_{Z|X}(z_1|x_1) \;=\; \frac{\delta\alpha\lambda + \gamma\beta(1-\lambda)}{\alpha\lambda + \beta(1-\lambda)} \tag{63}$$

Since the marginal probability that $Z = z_1$ is simply given by $p_Z(z_1) = p_{Z|Y}(z_1|y_1) \cdot p_Y(y_1) + p_{Z|Y}(z_1|y_2) \cdot p_Y(y_2)$, which given the above parameterization is expressed as $+\gamma(1-\lambda)$, the only way that the instantiated value of X could *fail* to affect the distribution of Z in this topology would be if

$$\frac{\delta\alpha\lambda + \gamma\beta(1-\lambda)}{\alpha\lambda + \beta(1-\lambda)} = \delta\lambda + \gamma(1-\lambda) \tag{64}$$

Clearing the denominator, expanding the resulting expression, and re-factoring gives us:

$$0 = (\alpha - \beta)(\delta - \gamma)(\lambda^2 - \lambda) \tag{65}$$

Therefore, the only way for one terminal variable in the 'fork' topology to *fail* to influence the other terminal variable would be either for the central variable Y to be instantiated to one of its specific values (which would make $\lambda^2 - \lambda = 0$ in the above expression), or for variables X and Y or variables Z and Y to be independent (which would make $\alpha = \beta$ or $\delta = \gamma$ respectively). *Hence once again the instantiation of the central variable blocks the influence of each terminal variable upon the other.*

APPENDIX C

D-SEPARATION WITHIN A COMMON EFFECT STRUCTURE

A ‘collider’ (or common effect) connection X → Y ← Z consists of three random variables *X*, *Y*, and *Z* for which both *X* and *Z* each directly influence variable *Y*, but for which there is no *direct* influence between variables *X* and *Z*. (Due to the lack of a direct connection between variables *X* and *Z*, this structure is also commonly called an *unshielded collider* in the literature.) Given this graphical structure, the joint probability of the three random variables X, Y, and Z is given by:

$$p_{X,Y,Z} \;=\; p_{Y\,|\,X,Z} \cdot p_X \cdot p_Z \tag{66}$$

We will discuss this form of connection with regards to multinomial data only. Further details can be found in Koski and Noble (2009) and Koller and Friedman (2009)

Since $p_{Y|\,X,Z} = p_{X,Y,Z}\,/p_{X,Z}$ for any three variables X, Y, Z having $p_{X,Z} \neq 0$, we can rewrite the above topology-specific factorization as:

$$p_{X,Y,Z} \;=\; p_X \cdot p_Z \cdot \left(\frac{p_{X,Y,Z}}{p_{X,Z}}\right) \tag{67}$$

Hence we have that $p_{X,Y,Z} \cdot p_{X,Z} \;=\; p_X \cdot p_Z \cdot p_{X,Y,Z}$, so as long as $p_{X,Y,Z} \neq 0$ we can reduce this expression to $p_{X,Z} = p_X \cdot p_Z$. Therefore, as long as both $p_{X,Y,Z}$ and $p_{X,Z}$ are nonzero, we have that *the two terminal variables in a collider connection are unconditionally independent of each other*. In other words, *contrary to the situation we had earlier* with the chain and fork connections, communication between the terminal variables *X* and *Z* in the collider connection is *blocked* when the intervening variable *Y* is *uninstantiated*.

We will now show that (once again contrary to what was the case with the chain and fork connections), the *instantiation* of the central variable Y in this topology actually *creates* a link between the two terminal variables X and Z, and thereby facilitates communication between these two variables. Prior to the instantiation of Y, we can marginalize the joint distribution of X, Y, and Z over the values of Y, obtaining:

$$\begin{aligned} p_{X,Z}(X,Z) &= \sum_y p_{Y|X,Z}(y|X,Z)\, p_X(X)\, p_Z(Z) \\ &= p_X(X)\, p_Z(Z) \cdot \sum_y p_{Y|X,Z}(y|X,Z) \\ &= p_X(X) \cdot p_Z(Z) \end{aligned} \tag{68}$$

On the other hand, instantiation of the variable Y in this topology will result in the conditional distribution:

$$p_{X,Z|Y}(X,Z|y) = \frac{p_{Y|X,Z}(y|X,Z)\, p_X(X)\, p_Z(Z)}{\sum_x \sum_z p_{Y|X,Z}(y|x,z)\, p_X(x)\, p_Z(z)} \tag{69}$$

This expression is not a simple product of $p_{X|Y}(X|y)$ and $p_{Z|Y}(Z|y)$, meaning that once the variable Y is instantiated, variables X and Z are no longer independent. Therefore, instantiation of the central variable in the collider topology allows the two terminal variables in this connection communicate with each other.

Alternatively, one can solve for $p_{Z|X,Y}(Z|x,y)$ rather than $p_{X,Z|Y}(X,Z|y)$ in order to obtain a somewhat simpler expression showing the conditional dependence of variables X and Z when Y is instantiated. In this case, we are essentially assuming that we have already instantiated Y and that now we also want to propagate the additional instantiation of X through the collider connection. Thus we compute:

$$\begin{aligned} p_{Z|X,Y}(Z|x,y) \;&=\; \frac{p_{X,Y,Z}(x,y,Z)}{p_{X,Y}(x,y)} \;=\; \frac{p_{Y|X,Z}(y|x,Z)\, p_X(x) p_Z(Z)}{\sum_z \, p_{Y|X,Z}(y|x,z)\, p_X(x) p_Z(z)} \\ &=\; \frac{p_X(x)\,\cdot\, p_Z(Z)\, p_{Y|X,Z}(y|x,Z)}{p_X(x)\cdot \sum_z \, p_{Y|X,Z}(y|x,z)\, p_Z(z)} \\ &=\; \frac{p_Z(Z)\, p_{Y|X,Z}(y|x,Z)}{\sum_z \, p_{Y|X,Z}(y|x,z)\, p_Z(z)} \end{aligned} \tag{70}$$

Since this expression is not equal to $p_Z(z)$, we have that in the context of an instantiated central variable Y, the additional instantiation of X will affect the distribution of the other terminal variable Z in the collider connection. Instantiating Y has *enabled* communication between X and Z within the collider topology.

APPENDIX D

ALTERNATIVE CHARACTERIZATIONS OF D-SEPARATION

The Bayesian network representation of the set of joint relationships among a group of modeled variables facilitates the efficient manipulation of relevance and irrelevance relationships (i.e., relations of both marginal and conditional independence and dependence) among the modeled constructs. As Pearl (1988) states,

> The advantage of network representation is that it allows people to express directly the fundamental qualitative relationship of 'direct dependency'. The network then displays a consistent set of additional direct and indirect dependencies and preserves it as a stable part of the model, independent of the numerical estimates. (Pearl, 1988, p. 51)

The mechanism by which this is accomplished, i.e., the means by which the Bayesian network representation facilitates the efficient manipulation of independence and dependence relationships among variables, is termed the directed separation principle (or just 'd-separation' for short). Pearl's original characterization of the d-separation principle is given below :

> If $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ are three disjoint subsets of nodes in a DAG D, then $\boldsymbol{Z}$ is said to d-separate $\boldsymbol{X}$ from $\boldsymbol{Y}$, denoted $\langle\, \boldsymbol{X} \mid \boldsymbol{Z} \mid \boldsymbol{Y} \,\rangle_D$, if there is no path between a node in $\boldsymbol{X}$ and a node in $\boldsymbol{Y}$ along which the following two conditions hold: (1) every node with converging arrows is in $\boldsymbol{Z}$ or has a descendant in $\boldsymbol{Z}$, and (2) every other node is outside $\boldsymbol{Z}$. (Pearl, 1988, p. 117)

Another, more common formulation of the d-separation criterion is typically given in terms of the existence of structures known as *paths* and *blockers*. Following, e.g., Kjaerulff and Madsen (2008), we can state this alternative formulation of the d-separation criterion as follows:

> A path Φ between $\Phi = \langle u, \ldots, v \rangle$ in a DAG $\mathbb{G} = (\mathbf{V}, \mathbf{E})$ is said to be *blocked* by a subset $S \subseteq \boldsymbol{V}$ if Φ contains a vertex w such that either of the following conditions hold:
>
> 1. $w \in S$, and the edges of the path Φ ***do not*** meet head-to-head at w.
> 2. Neither w nor any of its descendants are in S, and the edges of the path Φ ***do*** meet head-to-head at w.

Rather than basing the d-separation criterion on a set of *negated* conditions (as in the definition from Pearl, 1988, given above), or on *partially negated* conditions (as in the definition by Kjaerulff and Madsen also given above), it is helpful to reframe d-separation as a *positively framed* (i.e., *non-negated*) set of conditions, as is done for instance in Korb and Nicholson (2004). Framed in this way, the d-separation criterion becomes:

> Given variables *X* and *Y*, along with a set of variables *Z* disjoint from both *X* and *Y*, the variables *X* and *Y* are d-separated given *Z* if and only if all paths Φ between *X* and *Y* are "cut" by one of the following graph-theoretic conditions:
>
> 1. Φ contains a chain $A \to B \to C$ such that $B \in Z$.
> 2. Φ contains a divergent connection $A \leftarrow B \to C$ such that $B \in Z$.
> 3. Φ contains an unshielded collider $A \to B \leftarrow C$ such that neither *B* nor any of *B*'s descendants is in *Z*.

Finally, it is very useful to point out that a highly practical *substitute* for the d-separation criterion does exist in the literature, namely the ***directed global Markov criterion*** of Lauritzen, Dawid, Larsen, and Leimer (1990), which is stated below:

> For a directed acyclic graph $\boldsymbol{G} = (\mathbf{V}, \mathbf{E})$ and disjoint subsets $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{S}$ of $\mathbf{V}$, any pair of vertices $a \in \boldsymbol{A}$ and $b \in \boldsymbol{B}$ will be d-separated by S whenever all paths from a to b are blocked by the vertices of $\boldsymbol{S}$ within the *moralized graph* consisting of the ancestors of $\boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{S}$, i.e., the graph $\left(\boldsymbol{G}_{An(A \cup B \cup S)}\right)^{M}$.

According to Kjaerulff & Madsen (2008), the directed global Markov condition is quite often much easier to apply than the traditional d-separation criterion. (In fact, many authors refer alternately to *Lauritzen d-separation* when referring to the directed global Markov condition, and *Pearl-Geiger d-separation* when referring to the original version.)

To illustrate the usefulness of the directed global Markov condition (i.e., of Lauritzen d-separation), Kjaerulff & Madsen (2008) provide the following illustrative example of a DAG along with three sets of variables $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{S}$, and ask whether it is true that $(\boldsymbol{A} \perp \boldsymbol{B} | \boldsymbol{S})_{\boldsymbol{G}}$ within this graphical structure (from Kjaerulff & Madsen, 2008):

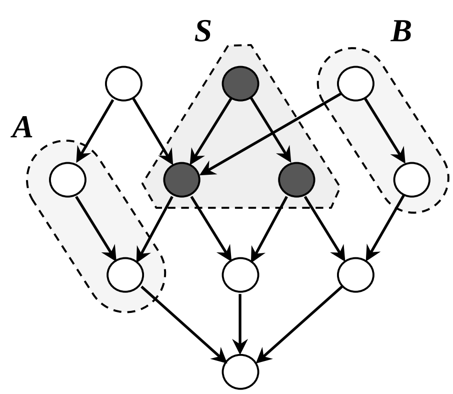


Figure 35. Structure representing a d-separation query

To utilize the directed global Markov condition, we reduce this graphical structure to just the *ancestral set* of $\boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{S}$ by removing any nodes that are not ancestors of at least some member of $\cup \boldsymbol{B} \cup \boldsymbol{S}$. This results in the reduced graphical structure $\boldsymbol{G}_{\mathrm{An}(\boldsymbol{A}\cup\boldsymbol{B}\cup\boldsymbol{S})}$ shown below (again, from Kjaerulff & Madsen, 2008):

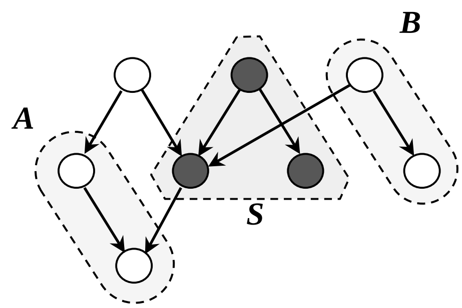


Figure 36. Ancestral set for a d-separation query

Next we *moralize* this structure by connecting (or 'marrying') the parents of any common children, and then replacing all directed links by undirected ones. This results in the structure $\left(\boldsymbol{G}_{An(A\cup B\cup S)}\right)^M$, from Kjaerulff & Madsen (2008):

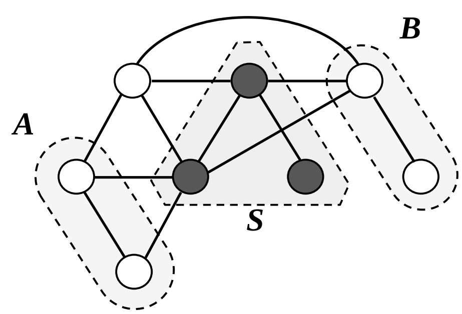


Figure 37. Moralized structure for a d-separation query

Now it is easy to determine whether or not set $\boldsymbol{S}$ d-separates sets $\boldsymbol{A}$ and $\boldsymbol{B}$. To do this, we merely need to check whether there is a path from $\boldsymbol{A}$ to $\boldsymbol{B}$ in this modified (moralized ancestral) graph that *avoids* passing through $\boldsymbol{S}$. In this example, there is such a path: namely the path from $\boldsymbol{A}$ to $\boldsymbol{B}$ involving the arc at the top of the structure. Therefore we easily conclude that $\boldsymbol{S}$ does <u>not</u> d-separate sets $\boldsymbol{A}$ and $\boldsymbol{B}$ in the original graph $\boldsymbol{G}$.