Strategic Management Journal
Strat. Mgmt. J., **35**: 1541–1551 (2014)
Published online EarlyView 8 August 2013 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2164
Received 23 December 2011; Final revision received 8 April 2013

RESEARCH NOTES AND COMMENTARIES

MODELS OF CAUSAL INFERENCE: IMPERFECT BUT APPLICABLE IS BETTER THAN PERFECT BUT INAPPLICABLE

FLORIAN ELLSAESSER,[1] ERIC W. K. TSANG,[2]* and JOCHEN RUNDE[3]
[1] *Front Row Society, Berlin, Germany*
[2] *Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas, U.S.A.*
[3] *Judge Business School, University of Cambridge, Cambridge, UK*

We assess a recent paper by Durand and Vaara (2009) that advances causal graph modeling as a tool for inferring causes in strategy research. We focus on the Markov condition, a key assumption on which causal graph modeling is based, and show why this condition is invariably violated in strategic management in general and the resource-based view of the firm in particular. We then introduce vector space modeling as a quantitative alternative to causal graph modeling, and consider how improved methods of causal inference might enhance our ability to test some of the central propositions of the resource-based view. Copyright © 2013 John Wiley & Sons, Ltd.

INTRODUCTION

Durand and Vaara (2009) maintain that an improved understanding of causation is essential for advancing strategy research and the resource-based view in particular. To this end, they define the epistemological conditions that lie at the foundation of causal inference and introduce a counterfactual conceptualization of causation. Based on this conceptualization, they then introduce two methodologies to infer causes, one qualitative (counterfactual history) and the other quantitative (acyclic causal graph modeling).

We applaud Durand and Vaara for their intervention. Causation is a complex issue that lies at the heart of strategy research (de Rond and Thietart, 2007). This is so, moreover, not only in relation to the methodologies employed by researchers, but also, in some cases, in relation to the practices of competing firms. Thus, while science is often seen as being about the provision of causal explanations in response to "why?" questions (Hempel and Oppenheim, 1948), in the world of business, according to the resource-based view, causal ambiguity may be an effective barrier to imitation and so help sustain competitive advantage (Barney, 1991; Lippman and Rumelt, 1982; Reed and DeFillippi, 1990). In this light, a more in-depth understanding of causation and how causes might be inferred could be a key ingredient for competitive success.

While Durand and Vaara see causal mechanisms as a useful way to conceptualize causal relationships, they also emphasize that one should not become bogged down in endless debates and excessive research conventions that get in the way of empirical research. We fully agree with this

Keywords: causation; causal graph modeling; vector space modeling; resource-based view; competitive advantage
*Correspondence to: Eric W. K. Tsang, Naveen Jindal School of Management, 800 West Campbell Road, SM 43, University of Texas at Dallas, Richardson, TX 75080-3021, U.S.A. E-mail: ewktsang@utdallas.edu

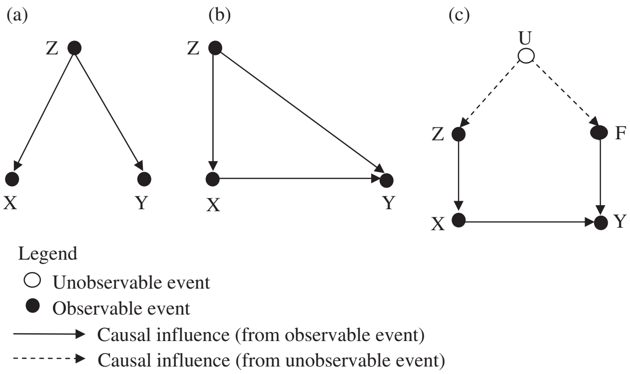

Figure 1. Back-door causal paths.
Notes: 1. Figures 1b and 1c are from Figures 2b and 2c, respectively, of Durand and Vaara (2009). 2. Figure 2b of Durand and Vaara on page 1258 is inconsistent. Z is shown as an unobservable event, but the causal arrows are not dotted. We thus take Z to be an observable event here

sentiment and that the ultimate aim of the kind of investigation they are engaged in should be to advance management theory in a way that has practical implications for businesses.

We do, however, have concerns about parts of their argument, and in particular with their treatment of causal graph modeling as a quantitative method for inferring causes. Durand and Vaara introduce three general varieties of this approach—conditioning, instrumenting, and mediating. They focus on mediating as the most promising of these three strategies (p. 1259, figure 4). Conditioning and instrumenting are not suitable for the resource-based view, as both require causal mechanisms that can be isolated. Our first objective is to show that causal graph modeling is also problematic for the strategy of mediating because the Markov condition on which it is based rarely holds in typical management situations. Given that our interest is in exploring quantitative approaches to causal analysis that may be useful in strategic management, we then consider vector space modeling as a possible alternative to causal graph modeling. The weight of attention here shifts from establishing whether one of a pregiven list of events lies in the causal history of some other event to uncovering causal mechanisms in situations of causal complexity.

THE FOUNDATION OF ACYCLIC CAUSAL GRAPH MODELING

The attraction of causal graph modeling is that, under certain conditions, it permits the determination of the causal relationships between types of events with logical necessity. Both Spirtes, Glymour, and Schein (1993) and Pearl (2000) prove theorems that show how causes can be identified through the probabilistic analyses of causal graph modeling.

As Durand and Vaara (2009: 1257) observe, "the general principle of causal graph estimation is to eliminate 'back-door paths' ...." A back-door path is constituted by any causal factor that influences the phenomenon to be explained through intermediate causes. Figure 1(a) illustrates a case in which there are no intermediate causes and therefore no back-door paths. Figure 1(b, c), which are equivalent to Durand and Vaara's Figure 2(b, c), illustrate cases in which the influence of Z on Y is felt via a back-door path through X.

The logic of blocking a back-door path is based on the logic of "screening-off" introduced by Reichenbach (1956). Formally, if events X and Y are statistically independent, they display the following relationship:

$$P\ (X\&Y) = P(X) \times P(Y) \qquad \text{(statistical independence)}$$

Events are statistically dependent if they occur together either more or less frequently than they would be expected to do by chance. This is the case when

$$P\ (X\&Y) \neq P(X) \times P(Y) \qquad \text{(statistical dependence)}$$

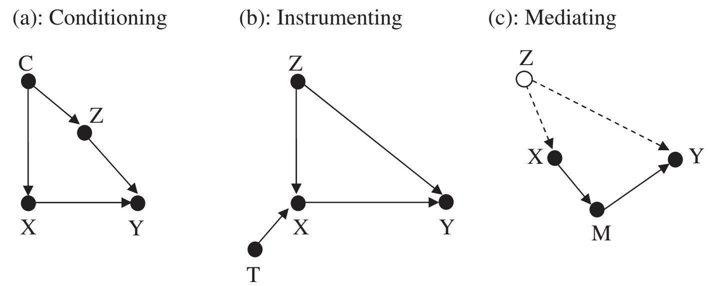

Figure 2. Three strategies for causal effect estimation.
Note: Figures 2b and 2c are from Figures 3b and 3c, respectively, of Durand and Vaara (2009)

Screening-off occurs when there is a common cause of two events that are initially statistically dependent.

Suppose there are three events X, Y, and Z, and that only the probabilistic relationships between them are known (see Figure 1a). Initially, there is a statistical dependence between events X and Y. Screening-off arises if there is a third event Z, which screens off the statistical dependence between the first two events X and Y. Given the third event Z, X and Y become statistically independent; that is:

$$\left.\begin{array}{l} P(X\&Y) \neq P(X) \times P(Y) \text{ (necessary condition)} \\ P(X\&Y/Z) = P(X/Z) \times P(Y/Z) \text{ (necessary condition)} \end{array}\right\} \text{Jointly sufficient screening-off conditions}$$

Consequently Z must lie in the causal history of X and Y.

The logic of screening-off can be extended to eliminate any back-door path. Take the situation depicted in Figure 1(c). Z influences Y through a back-door path. Hence, initially there is a statistical dependence between Z and Y:

$$P(Z\&Y) \neq P(Z) \times P(Y)$$

However, once we condition for X and F, Z becomes statistically independent of Y:

$$P(Z\&Y/X\&F) = P(Z/X\&F) \times P(Y/X\&F)$$

Consequently Z must lie in the causal history of either F, X, or both.

The first two strategies introduced by Durand and Vaara, conditioning (Figure 2a) and instrumenting (Figure 2b), are based on the back-door criterion. Conditioning involves accounting for all back-door paths ($C_1 \ldots C_n$) of one or more potential causal factors (X, Z). As shown in Figure 2(a), C becomes independent of Y after taking into account X and Z. Hence, one can establish that there is a causal chain from the back-door factor C through X and Z to Y. In the case of instrumenting, there is a controllable instrument (T) that directly influences X (Figure 2b). If T becomes independent of Y, given X, we know that there must be a causal chain from T through X to Y. Conditioning and instrumenting are thus both cases of the back-door path condition. While we are searching for a whole host of factors in the former, in the latter we already know of a factor T that indirectly triggers Y and with which the effect estimation can be carried out.

The third strategy presented by Durand and Vaara, mediating (Figure 2c), is based on the front-door condition. The front-door condition involves looking for a factor M that lies between the potential causal factor of interest X, and the outcome Y (Pearl, 2000). The front-door condition is fulfilled if X becomes statistically independent of Y again, once M is conditioned for. Such conditional independence implies that X causes M, which in turn causes Y.

Both the front- and back-door criteria and thus all three strategies for causal effect estimation are based on the logic of the screening-off. Events that were initially statistically dependent become statistically independent, given the conditioning for some other events. This conditional independence indicates that the events must lie in the causal history of some event under consideration. Conditioning and instrumenting use the back-door criterion, since there is a movement backwards in the causal chain to establish causal relationships. Mediating through M (as shown in Figure 2c) uses the front-door criterion, since a factor M that is a descendant of X is used to establish that X is causally relevant for Y. Durand and Vaara drop conditioning and instrumenting, as the assumptions on which they are based are often violated under real world conditions.

The fact that causal graph modeling is based on the logic of screening-off highlights one of the core assumptions on which it rests—the Markov condition. This condition states that, once all of its parents are conditioned for, a variable will be probabilistically independent of all other variables except for its descendants. It is important to recognize how fundamental this assumption is to Durand and Vaara's project, and how quickly the arguments favoring causal graph theory break down once it is violated.

LIMITATIONS OF ACYCLIC CAUSAL GRAPH MODELING

The main limitation of acyclic causal graph modeling, and the one we want to highlight here, is that the Markov condition frequently fails to hold under the kind of circumstances faced in management decision making. We consider two cases in which this is so.

Missing information

The Markov condition may fail to hold when there is insufficient information about potential causes and their effects. This includes cases in which not all relevant parents are specified and in which events have not been specified in a sufficiently fine-grained way (Arntzenius, 1992).

The problem of missing information can be illustrated using an adapted version of Durand and Vaara's prime example of the mediating strategy. Consider Figure 3, which is a simplified version of their Figure 4. In addition to excluding some variables, the main modification we have made is to place the unobservable event U in a slightly different position.

Suppose we want to establish whether certain resources R lead to high firm performance Y_t.

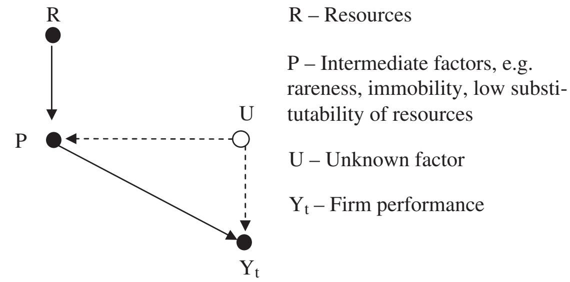

Figure 3. Missing information on causally relevant events

Assume that the individual and combined probabilities of these factors are known, and that R is mediated by P (a set of intermediate factors such as rareness, immobility, and low substitutability of resources). To be able to establish that Y_t is causally related to R in the way that Durand and Vaara propose, the following relations would have to hold:

$$P(R) \times P(Y_t) \neq P(R\&Y_t)$$

$$P(R/P) \times P(Y_t/P) = P(R\&Y_t/P)$$

In this case, R and Y_t would initially be statistically dependent but become statistically independent once we condition for P. We would then establish that R influences firm performance Y_t through some mediating factor P.

However, given the relationships depicted in our example, the following probabilities actually hold:

$$P(R) \times P(Y_t) \neq P(R\&Y_t)$$

$$P(R/P) \times P(Y_t/P) \neq P(R\&Y_t/P)$$

P does not make R and Y_t statistically independent because there is an unknown factor U, say whether the technological environment of an industry is changing rapidly, that directly influences P. Durand and Vaara accordingly have to assume that "the unobservable factors … do not influence P" (p. 1259) for their approach to work. The presence of any such unknown factors that directly influence either the mediating factor P or the potential causal factor R, and the phenomenon under consideration Y_t, will lead to the breakdown of the Markov condition, rendering causal graph modeling infeasible (Arntzenius, 1992; Salmon, 1985).

So while the Markov condition would hold in Durand and Vaara's example, if the relationship between R and P is not affected by any unknown factor U, this is highly unlikely ever to be the case. Most readers, for instance, would probably already have found our example overly simplistic. In real world situations,

(1) competitive advantage is likely to depend on a complex web of causal relationships with "many back-door paths," as Durand and Vaara (p. 1259) themselves argue.
(2) difficult-to-imitate resources tend to be associated with causal ambiguity. While the term causal ambiguity is a contested one (Powell, 2001), there is common agreement that it refers to some sort of complexity in the relevant situation and difficulties in codifying knowledge (King and Zeithaml, 2001; Reed and DeFillippi, 1990). Causal ambiguity thus implies that the generation of competitive advantage involves a multitude of factors (Ryall, 2009) and that some of these will be unknown.

In short, the Markov condition is unlikely to hold in the context of the resource-based view.

Interdependence and cyclical relationships

The previous section showed why the Markov condition may fail to hold because of a lack of relevant information. We now show that the Markov condition is unlikely to hold even under conditions of perfect information.

Consider again the Markov condition: *Once we condition for all its parents, a variable has to be probabilistically independent of all other variables, except its descendants*. To illustrate this requirement, take again the case of a parent Z that causes X and Y, as represented by Figure 1(a). For the Markov condition to hold, X and Y must be statistically independent, given that Z has occurred. This would be equivalent to a situation in which, once Z has occurred, the likelihood of X and Y is determined by tossing two fair coins independently of each other. However, it seems to be much more likely intuitively that the effects of X and Y will depend on each other in some way, given that they issue from a common cause (Cartwright, 1999).

Consider a more concrete example: Suppose a firm purchases a new machine (Z) that will, on average, reduce the defect rate of output (X) and the chance of machine breakdown (Y), as compared with the old machine. The Markov condition demands nothing less than that, once the purchase of the new machine is taken into account, reductions in the defect rate and reductions in machine breakdown will occur independently of each other, as if the occurrence of each type of events were determined by tossing a coin. Yet, since both types of events result from a common cause, it is surely reasonable to expect that their occurrences are correlated. For instance, when the machine is close to a breakdown state, the defect rate is likely to be higher. In cases like this, it is highly unlikely that the Markov condition will hold.

An alternative example runs as follows: It has long been accepted that firms exist precisely because they are complex, interdependent structures (Coase, 1937) in which the output of the combined work of employees is larger than the output of its constituent parts (Alchian and Demsetz, 1972). Similar synergies are observed for firms' external relations. Cohen and Levinthal (1990), for example, argue that firms that have prior knowledge in a particular area are better at acquiring new knowledge in the same area than firms without such prior knowledge. The newly acquired knowledge will contribute to the knowledge base that further enhances the acquisition of knowledge in that area. Knowledge interactions of this kind create a virtuous loop that can help firms to get ahead of the competition.

Such complexities are particularly important for the resource-based view. As Reed and DeFillippi (1990: 89) point out: "Although these works vary in their descriptions of competency, they remain consistent on two important themes: (1) the source of competency is always *internal* to the firm, (2) competency is produced by the way a firm utilizes its *internal* skills and resources, relative to the competition" (emphasis added). Hence, in a structure as complex as a firm, it is unlikely that the consequences of a common cause can be regarded as analogous to two independently tossed coins. Rather, since the effects result from the same parent, it is more likely that they will be interdependent, and in which case the Markov condition cannot be satisfied.

Durand and Vaara only acknowledge the lesser of these limitations of causal graph modeling: "It should be noted that causal graphs are

DOI: 10.1002/smj

non-parametric and acyclic (i.e., they do not permit representation of circular causation ...)" (p. 1257). They do not mention that causal graph modeling also requires that causes lead to their effects independently, a particular form of atomism according to which causal mechanisms exert their effect in isolation from each other. Such atomism is unlikely to hold in the present case, since it is in the very nature of firms to be composed of structured, interdependent relationships.

VECTOR SPACE MODELING AS AN ALTERNATIVE

If causal graph modeling is not generally applicable under the conditions that prevail in strategy research, are there any alternative quantitative techniques that may be of use? We suggest that techniques based on vector space modeling offer a promising possibility. In contrast to causal graph modeling, vector space-based algorithms have a large number of successful scientific and business applications in open systems in which causal mechanisms cannot be easily isolated. Functional applications include search engines (Berry and Young, 1995), literature-based discovery where previously unknown relationships between phenomena are inferred (Swanson, 1988), image recognition (Bulcão-Neto *et al.*, 2011), web-based translation (Bishop, 2006), and meaning recognition of language, where vector space-based models have been shown to outperform the average human on both TOEFL (Rapp, 2003) and SAT tests (Turney and Pantel, 2010). Vector space-based algorithms are also central to the success of a number of businesses:

- Google search, the fortunes of which are largely attributed to it being able to return the most relevant search results, is based on a vector space-based algorithm (Bishop, 2006).[1]
- Apple's iTunes, which uses a vector space-based algorithm to infer what customers like and buy based on their existing playlists (Christopher, 2012).
- Pandora Internet Radio, which tailors its online music channels to its customers in a way similar to iTunes and had a 2011 IPO valuation of US $2.6 billion (Fuller, 2006).

Vector space models such as latent semantic indexing or Dirichlet allocation were originally developed to identify similarities in linguistic concepts (Blei, Ng, and Jordan, 2003). Although the models themselves can be highly sophisticated, the underlying logic is straightforward. States of affairs are represented as points in a multidimensional space, with each aspect of a state of affairs occupying a particular dimension. The points are modeled as vectors that have a length and a direction (hence the term vector space modeling).

The working of the algorithms can be illustrated using the analogy of linguistic text, the application they were originally developed for. A text has two characteristics; namely, that it

(1) depicts a number of different entities, also called terms; and
(2) contains information on how the entities are structurally related.

In a piece of text, each word represents a basic unit of information. Words, in turn, are composed of letters. To infer the meaning of a word or string of words, vector space-based algorithms analyze how words and the letters they are composed of relate. By analyzing multiple texts, the meaning of words in their particular context can be identified. Texts do not have to be composed of the same words to have similar meanings. The algorithms are able to pick up similar structural arrangements, even if some of the words differ within the texts (Bishop, 2006).

Switching now to the case of the resource-based view, a firm's particular structural configuration might be regarded as analogous to text. This text comprises intra- and interorganizational processes (the words) such as particular routines, how these routines are internally structured (the letters that comprise the words), and how the routines relate (as words form sentences in a text).

Let us now consider an example. Suppose we hypothesize that the way a firm is integrated with its suppliers is a key structural feature in achieving competitive advantage in terms of R&D performance, and that it is difficult to imitate such integration because there is causal ambiguity as to whether and which elements of the relationship lead to the competitive advantage.

[1] Google's actual algorithm is a closely guarded secret, but it is widely known that modern search engines use vector space-based inference of one sort or another (Nastic, 2010; Wilson, 2005).

Suppose further that whether a desired type of relationship can be achieved depends on the wider culture in which the firm operates. For instance, building relationships of mutual trust might require substantial investment and might be difficult to achieve in a society with a "transactional" culture. We will now use this example to illustrate how vector space-based modeling differs from acyclic causal graph modeling in terms of (1) the type of data used, (2) the mechanism that converts the data input into an output, (3) the nature of the output, and (4) the assumptions made.

The type of data used

A vector space-based modeling exercise might take information about meetings between a company and its suppliers from computer-based diaries as a source of data. Such diaries typically contain names of participants, their companies, their rank within these companies, and the topic(s) of each meeting. Imagine that we want to investigate R&D performance. This might involve collecting key R&D metrics, such as which products were developed and how successful the products were in terms of indicators like speed of development, budget, and sales. While the researcher would still have to decide what data, such as the structure of meetings, he or she wants to collect, the algorithm does not require the inputting of a classification of the structures that these meetings take. Establishing such structures will rather be an outcome of the analysis.

In contrast, acyclic causal graph modeling would require the researcher to define a set of constructs and variables that measure the key factors he or she believes to be causally relevant. As we have seen in Durand and Vaara's example, these constructs frequently describe whether firms possess particular characteristics or resources. Examples of the types of categories that the firms under investigation might have to be slotted into might include whether they have a matrix or a pyramid structure, how frequently they hold joint meetings with their suppliers, the level of trust established with the suppliers, and whether they have joint development teams with their suppliers. Identifying clearly defined entities or properties in advance, rather than their structural relations, therefore becomes the main focus of data collection.

The mechanism that converts the data input into an output

In the vector space-based model, the composition of meetings is comparable to written text in the analysis of linguistic meaning. Each processual feature of a meeting, such as its participants, the departments they are affiliated with, and the companies they belong to, is depicted in multidimensional vector space. The resulting vector represents the total structural description of each meeting. The similarity between structures can then be determined by calculating the angles between the vectors that represent them, with lower angles of deviation indicating higher structural similarity. If, for example, the same departments are present in a number of meetings, these meetings will be considered structurally more similar in that respect.

In contrast, acyclic causal graph models assume perfect knowledge of the probabilities. This implies that traditional statistical analysis is required to (1) identify the conditional probabilities between the entities in question, and (2) assess whether the sample probabilities correspond to the population probabilities. For example, we might take 1,000 firms and determine whether they have an arm's length relationship with their suppliers or whether they form strategic alliances, and then look at the conditional probabilities of arm's length relationships in conjunction with R&D performance characteristics such as speed of product development.

The nature of the output

The way in which the data are converted into an output influences the nature of that output. In the case of the vector space-based model, the output could be a characterization of the type and structure of interaction that is associated with the development of particularly successful products. Possible findings might include that

- a particular structure of interaction (e.g., between different departments) is fruitful,
- a particular structural evolution of interactions over time is fruitful (e.g., first interaction between certain departments of the focal firm and the supplier and then interaction between some of the firm's own departments),
- some structures of interaction are more common if strategic alliances are present, or that

DOI: 10.1002/smj

- particular compositions of teams may be effective.

Causal graph modeling, in contrast, would require population-based probabilities and that the Markov condition holds. The causal graph model then structures the conditional probabilities in terms of causal relationships. We may then find, for example, that a long-term relationship with suppliers, such as a joint venture, lies in the causal history of successful product developments.

The assumptions made

In conclusion, then, the differences between vector space modeling and causal graph modeling boil down to the assumptions made. Vector space modeling combines the inference of structural relations with the inference of causal directionality as two steps of an inseparable problem. Thus, it identifies causal structures where the relata have an internal structure and interdependencies between relata drive the competitive advantage of firms. The aim is to provide mechanismic explanations (Bunge, 1997), which describe the causal mechanisms underlying the phenomena concerned. Some of these structures may share similarities, and, if so, it might then be possible to identify the types of structural relations associated with competitive advantage. We can name and describe these structures, but no two will be exactly the same, as each varies in its elements and their arrangements. As such, vector space modeling is particularly well suited to situations that involve a complex web of causation where no single factor, but an interdependent web of causes, leads to competitive advantage. Uncovering the underlying causal mechanisms can be of considerable help in understanding these situations.

In contrast, causal graph modeling assumes perfect knowledge of the probabilistic relationship between events and that the main task is to draw conclusions about causal relationships from these probabilities. Hence, it represents a view of event causation (Lewis, 1973) where clearly definable, identifiable, and separable entities exist, which we call resources. In particular, it is assumed that the influence of one entity is clearly directed at another, so that there are no mutual interdependencies, and that entities can be measured and their causal influence separated. The effects of the entities are assumed to be conditionally independent, and therefore it is assumed that the Markov condition holds. Further, while entities may causally relate to each other, they are denied any internal structure, which, according to vector space modeling, is crucial for generating causal mechanisms. Table 1 summarizes the comparison between vector space modeling and causal graph modeling.

BENEFITS AND LIMITATIONS OF VECTOR SPACE MODELING

Perhaps the most important benefit of vector space modeling is that it can deal with the causal complexity that is associated with sustained competitive advantage. As argued above, vector space models do not depend on the Markov condition, which breaks down under the conditions stated by the resource-based view. Even if the Markov condition were to hold, acyclic causal graph modeling would not give us results that sufficiently take account of causal complexity. Coming back to our example of what causes superior R&D performance, for instance, a causal graph model may return that a pyramid structure is more effective than a matrix structure, or that cross-functional meetings are more effective than single-function meetings. These results are very limited, as they reduce causation to a few, supposedly independent factors. In contrast, vector space models give us an understanding of the underlying mechanism and consequently provide insights into the complex web of causation that typically leads to competitive advantage. A vector space model may, for example, identify that joint product development teams between a company and its supplier, where a variety of ranks meet frequently inside and outside of work to create a high trust culture, lead to joint R&D success.

How does vector space modeling perform when applied to a complex web of causal interactions and missing data? Suppose the researcher had access to the transcripts of R&D meetings but did not have access to other rich sources, such as R&D expenditures by project and scientist locations. It turns out that vector space modeling outperforms methods that require traditional statistical analysis, such as acyclic causal graph modeling, in cases of causal complexity and missing data (Duch, Swaminathan, and Meller, 2007). Even if we had only the transcripts of R&D meetings, these documents will contain numerous potential

DOI: 10.1002/smj

Table 1. Comparison between causal graph modeling and vector space modeling

	Causal graph modeling	Vector space modeling
1. Nature of data	• Clearly defined and separable constructs • Firms need to be classified in terms of the constructs of interest	• Main focus on structural elements and their relations • Less preclassification necessary
2. Mechanism that generates output	• Inference of population statistics • Conditional independence	• Comparing deviation of angles between vectors • Identifying structural similarities (usually by calculating cosine between vectors)
3. Nature of output	• Causal relations of various significance levels and strengths between constructs	• Structural similarities between objects or features of objects such as firms
4. Assumptions	• Perfect knowledge of probabilistic relationships between events • Separable, independently acting resources drive competitive advantage (event causation) • Causal effects are conditionally independent (Markov condition)	• Structural characteristics of firms drive competitive advantage (structural causation) • Key to these structural characteristics are the relations in which entities stand to each other • Resources are mutually dependent entities that cannot be entirely isolated from their context

causal factors. Thus, an almost infinite number of potential combinations could explain the effect, leaving the internal structure of the relevant mechanism a black box. It is highly likely that methods such as multivariate regressions or genetic algorithms will result in overfitting and the identification of spurious relationships. As causal graph modeling requires a similar type of data input, it suffers from the same problem. Vector space modeling, on the other hand, provides a sense of the internal structure of a mechanism, which is composed of a multiplicity of components as one vector. It is thus much less likely that spurious correlations are identified as

(1) the number of potentially explanatory factors is substantially lower, and
(2) we obtain a sense of the internal structure of the complex web of causation and can thus verify whether the causal mechanism, so identified, is plausible.

There is clear evidence for the advantages of vector space modeling in pharmaceutical research. This is an area in which significant advances have been made over the last thirty years by focusing increasingly on the mechanisms and pathways through which drugs work, rather than on merely whether a drug is efficacious and safe (Rainsford, 1995). Vector space modeling has been shown to be more effective than traditional statistical approaches in this context, as it takes into account a variety of complex structures such as the three-dimensional structure of molecules and the interaction of a number of different genes (Nobel, 2006). The superiority of vector space modeling has also been confirmed by a number of direct comparisons with alternative techniques, such as neural networks: Vector space-based models have the highest reliability for a number of tasks, such as structure-activity analysis (Burbidge *et al.*, 2001), drug/nondrug classification (Byvatov *et al.*, 2003), intestinal absorption (Hou, Wang, and Li, 2007), and agrochemical similarities (Zernov *et al.*, 2003).

Of course, vector space-based models have limitations, too. In the first place, they require detailed data on the underlying structures and processes. Second, it may be difficult to acquire such data from firms and to transform the data into a format that can be used by the algorithms. And third, sometimes there may be only few observations of key variables. For example, if a critical decision leading to R&D success is made in a single meeting and is not captured by the data, vector space modeling will not be able to identify this causal factor. Of course, causal graph modeling will fail too in this instance, as no correlations can be identified. A more appropriate alternative here could be contrastive explanation, which focuses on answering questions of the form "why X rather than Y?" where X is the fact to be explained and Y the foil. Causes can then be inferred by comparing the fact, which is the event of interest, with a foil, which has a similar

DOI: 10.1002/smj

causal history but did not lead to the event of interest (Tsang and Ellsaesser, 2011). The idea is that potential causes are likely to be located where the causal histories of the fact and the foil differ (Lipton, 1991). In brief, constrastive explanation allows us to identify causally relevant factors by comparing the situation with an alternative path of action (Runde and de Rond, 2010). Although contrastive explanation cannot solve the problem of missing a crucial causal factor (and no method can), it generally performs better than other quantitative methods in situations where researchers can only have few observations of key variables.

A fourth limitation is that we need to select the structures to be studied, such as the meetings discussed in our example above. So there is always some judgment involved. However, in contrast to causal graph modeling, it is not necessary to define in advance which particular properties of a meeting, such as their cross-functional nature, could be the cause of competitive advantage, and to classify the meetings accordingly. Rather, such structural characteristics will emerge from the analysis.

Finally, while vector space modeling returns information that helps to uncover underlying causal mechanisms, it cannot guarantee that the mechanism identified was causally relevant in any particular case. Making causal attributions ultimately involves an unavoidable element of judgment, which will depend also on the background knowledge, experience, and intuition of the investigator concerned.

CONCLUSION

Durand and Vaara (2009) deserve credit for highlighting the importance of causation in strategy research and for introducing acyclic causal graph modeling as an approach to studying causal relationships. However, despite offering a glimpse of a perfect world in which causes can be inferred with logical necessity, causal graph modeling faces serious limitations. We have highlighted its dependence on the Markov condition, which will seldom hold in strategic management, its presupposing a world of events or states of affairs whose internal structures cannot be accounted for, and its reliance on perfect knowledge of probabilistic relationships.

Vector space-based modeling operates on the premise that it is usually structures and structural relations that lead to competitive advantage. While we maintain that vector space modeling may be of considerable help in identifying structures and processes that may be a cause of competitive advantage, we recognize that it is imperfect insofar as it cannot identify causal relationships with logical necessity.

From a pragmatic point of view, we suggest that imperfect but applicable is better than perfect but inapplicable, and that the extent of success in real world applications should be the ultimate standard of judgment. Even if it is not possible to fully establish causality, research can still be useful as long as it is making progress towards that end. Acyclic causal graph modeling has shown some success in reproducing already-known causal relationships in molecular biology under experimental conditions in which it is possible to be reasonably certain that the Markov condition holds. But these are very special cases. Vector space-based models, in contrast, already have a very good track record in open, nonexperimental conditions.

ACKNOWLEDGEMENT

We thank Richard Bettis (co-editor) and two reviewers for their valuable comments.

REFERENCES

Alchian AA, Demsetz H. 1972. Production, information cost, and economic organization. *American Economic Review* **62**(5): 777–795.

Arntzenius F. 1992. The common cause principle. *Proceedings of the Biennial Meeting of the Philosophy of Science Association* **2**: 227–237.

Barney JB. 1991. Firm resources and sustained competitive advantage. *Journal of Management* **17**(1): 99–120.

Berry MW, Young PG. 1995. Using latent semantic indexing for multilanguage information retrieval. *Computers and the Humanities* **29**(6): 413–429.

Bishop CM. 2006. *Pattern Recognition and Machine Learning*. Springer: New York.

Blei DM, Ng AY, Jordan MI. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* **3**: 993–1022.

Bulcão-Neto RF, Camacho-Guerrero JA, Dutra M, Barreiro A, Parapar J, Macedo AA. 2011. The use of latent semantic indexing to mitigate OCR effects of related document images. *Jounral of Universal Computer Science* **17**(1): 64–80.

Bunge M. 1997. Mechanism and explanation. *Philosophy of the Social Sciences* **27**(4): 410–465.

Burbidge R, Trotter M, Buxton S, Holden S. 2001. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Computer and Chemistry* **26**(1): 5–14.

Byvatov E, Fechner U, Sadowski J, Schneider G. 2003. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Journal of Chemical Information and Computer Sciences* **43**(6): 1882–1889.

Cartwright N. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press: Cambridge, UK.

Christopher M. 2012. How iTunes genius really works. *Technology Review*. Available at: http://www.technologyreview.com/view/419198/how-itunes-genius-really-works (accessed 23 June 2012).

Coase RH. 1937. The nature of the firm. *Economica* **4**(16): 386–404.

Cohen WM, Levinthal DA. 1990. Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly* **35**(1): 128–152.

Duch W, Swaminathan K, Meller J. 2007. Artificial intelligence approaches for rational drug design and discovery. *Current Pharmaceutical Design* **13**(14): 1497–1508.

Durand R, Vaara E. 2009. Causation, counterfactuals and competitive advantage. *Strategic Management Journal* **30**(12): 1245–1266.

Fuller B. 2006. Inside Pandora: web radio that listens to you. *O'Reilly Digital Media*. Available at: http://openp2p.com/pub/a/oreilly/digitalmedia/2006/08/17/inside-pandora-web-radio.html (accessed 23 June 2012).

Hempel CG, Oppenheim P. 1948. Studies in the logic of explanation. *Philosophy of Science* **15**(2): 135–175.

Hou T, Wang J, Li Y. 2007. ADME evaluation in drug discovery. 8. The prediction of human intestinal absorption by a support vector machine. *Journal of Chemical Information and Modeling* **47**(6): 2408–2415.

King AW, Zeithaml CP. 2001. Competencies and firm performance: examining the causal ambiguity paradox. *Strategic Management Journal* **22**(1): 75–99.

Lewis D. 1973. Causation. *Journal of Philosophy* **70**(17): 556–567.

Lippman SA, Rumelt RP. 1982. Uncertain imitability: an analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics* **13**(2): 418–438.

Lipton P. 1991. Contrastive explanation and causal triangulation. *Philosophy of Science* **58**(4): 687–697.

Nastic R. 2010. What is latent semantic indexing? *Search Engine Journal*. Available at: http://www.searchenginejournal.com/what-is-latent-semantic-indexing-seo-defined/21642 (accessed 26 August 2012).

Nobel WS. 2006. What is a support vector machine? *Nature Biotechnology* **24**(12): 1565–1567.

Pearl J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press: Cambridge, UK.

Powell TC. 2001. Competitive advantage: logical and philosophical considerations. *Strategic Management Journal* **22**(9): 875–888.

Rainsford KD (ed)., 1995. *Progress in Drug Research*. Springer: Basel, Switzerland.

Rapp R. 2003. Word sense discovery based on sense descriptor dissimilarity. Proceedings of the Ninth Machine Translation Summit, New Orleans, LA; 315–322.

Reed R, DeFillippi RJ. 1990. Causal ambiguity, barriers to imitation, and sustainable competitive advantage. *Academy of Management Review* **15**(1): 88–102.

Reichenbach H. 1956. *The Direction of Time*. University of California Press: Berkeley, CA.

de Rond M, Thietart RA. 2007. Choice, chance, and inevitability in strategy. *Strategic Management Journal* **25**(8): 535–551.

Runde J, de Rond M. 2010. Evaluating causal explanations of specific events. *Organization Studies* **31**(4): 431–450.

Ryall MD. 2009. Causal ambiguity, complexity, and capability-based advantage. *Management Science* **55**(3): 389–403.

Salmon W. 1985. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press: Princeton, NJ.

Spirtes P, Glymour C, Schein R. 1993. *Causation, Prediction and Search*. Springer: New York.

Swanson D. 1988. Migraine and magnesium: eleven neglected connections. *Perspectives in Biology and Medicine* **31**(4): 526–557.

Tsang EWK, Ellsaesser F. 2011. How contrastive explanation facilitates theory building. *Academy of Management Review* **36**(2): 404–419.

Turney PD, Pantel P. 2010. From frequency to meaning: vector space models of semantics. *Journal of Artificial Intelligence Research* **37**: 141–188.

Wilson R. 2005. How does Google use latent semantic indexing (LSI)? *Web Marketing Today*. Available at: http://webmarketingtoday.com/articles/google-lsi (accessed 26 August 2012).

Zernov VV, Balakin KV, Ivashenko AA, Savchuk NP, Pletnev IV. 2003. Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical likeness, and enzyme inhebition predictions. *Journal of Chemical Information and Computer Sciences* **43**(6): 2048–2056.