University of Toronto

From the SelectedWorks of Michael D Ryall

March, 2009

Causal Ambiguity as a Source of Sustained Capability-Based Advantages

Michael D Ryall, *University of Toronto*

Available at: https://works.bepress.com/michael_ryall/10/

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

MANAGEMENT SCIENCE
Vol. 55, No. 3, March 2009, pp. 389–403
ISSN 0025-1909 | EISSN 1526-5501 | 09 | 5503 | 0389

informs®

DOI 10.1287/mnsc.1080.0938
© 2009 INFORMS

Causal Ambiguity, Complexity, and Capability-Based Advantage

Michael D. Ryall
University of Melbourne, Carlton, Victoria 3053, Australia, mdr@mikeryall.com

This paper presents the first formal examination of role of causal ambiguity as a barrier to imitation. Here, the aspiring imitator faces a knowledge (i.e., "capabilities-based") barrier to imitation that is both causal and ambiguous in a precise sense of both words. Imitation conforms to a well-explicated process of learning by observing. I provide a precise distinction between the intrinsic causal ambiguity associated with a particular strategy and the subjective ambiguity perceived by a challenger. I find that intrinsic ambiguity is a necessary but insufficient condition for a sustained capability-based advantage. I also demonstrate that combinatorial complexity, a phenomenon that has attracted the recent attention of strategy theorists, and causal ambiguity are distinct barriers to imitation. The former acts as a barrier to explorative/active learning and the latter as one to absorptive/passive learning. One implication of this is that learning by doing and learning by observing are complementary strategic activities, not substitutes—in most cases, we should expect firm strategies to seek performance enhancement using efforts of both types.

Key words: causal ambiguity; competitive inference; *NK* complexity; sustained advantage; imitation barrier; learning by observing; learning by doing
History: Accepted by Pankaj Ghemawat, business strategy; received March 24, 2004. This paper was with the author 1 year and 5 months for 2 revisions. Published online in *Articles in Advance* November 18, 2008.

1. Introduction

A central proposition in strategy is that firms sustain relative performance advantages only if their existing and potential rivals cannot imitate them (Nelson and Winter 1982, Dierickx and Cool 1989, Barney 1991). In this context, "imitation" means the purposeful endeavor to improve performance by copying the form and strategy of a superior rival. An imitation strategy is one of many ways two firms may become similar in appearance and performance. For example, de novo innovation can result in such similarities and, when it does, also be referred to as "imitation" (as in Lippman and Rumelt 1982). This paper is the first to provide a formal analysis of causal ambiguity as a barrier to imitation.

Generally speaking, imitation fails when it is physically impossible, legally prevented, economically unattractive, or the necessary knowledge is lacking. Saloner et al. (2001) label barriers of the first three types "positional" and those of the last "capabilities based." The conditions leading to positional barriers (e.g., switching costs, entry costs, scope and scale economies, and the likelihood of ex post retaliation) have been extensively studied in both formal and informal settings and are presently well understood (e.g., Porter 1980, Tirole 1988).

Capabilities-based barriers have received much less in the way of formal attention. When imitation is hampered by a lack of knowledge, learning becomes a central issue. Learning can be explorative in the active sense of learning from one's own experience (learning by doing), or absorptive in the passive sense of learning from external information (learning by observing).[1] Thus, a capabilities-based advantage is sustained only if learning of both types is effectively blocked.

An important exception to the scarcity of formal work in this area is the recent stream examining combinatorial complexity as a capabilities-based barrier to imitation (Levinthal 1997; Ghemawat and Levinthal 2000; Rivkin 2000; Lenox et al. 2006, 2007). This work applies the formalism of Kaufman's (1993) "*NK*" model of evolutionary biology to competition between firms. In this setting, managers do not know the relationship between activities and performance. They learn by exploring in the "local neighborhood" of their current activities or, if attempting to imitate,

[1] There are multiple streams of literature in strategy that focus on learning by doing, each deserving a survey paper in its own right. These include papers on learning curves (e.g., Lieberman 1984, Ghemawat 1985), exploration (e.g., March 1991, Levinthal and March 1993), dynamic experimentation (e.g., Besanko et al. 2007), and search (e.g., Lippman and Rumelt 1982, Rivkin 2000). Learning by observing has seen much less attention, a notable exception being the work on absorbtive capacity (e.g., Cohen and Levinthal 1990).

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

in the neighborhood of an industry leader. When combinatorial complexity is high, interactions between activities make local exploration much less effective. Thus, combinatorial complexity can be a source of sustained capabilities-based advantage.

The focus here is on a different obstacle to imitation, causal ambiguity. In my analysis, causal ambiguity operates as a barrier to absorptive, or passive, learning. The term "causal ambiguity" in its traditional usage refers to any knowledge-based impediment to imitation (e.g., Saloner et al. 2001, p. 49). The first strategy paper using this term appears to be Lippman and Rumelt (1982, p. 418), who assert, "basic ambiguity concerning the nature of the causal connections between actions and results" can result in persistent performance heterogeneity because "the factors responsible for performance differentials resist precise identification." Although Lippman and Rumelt (1982) present a formal model, causal ambiguity does not enter into it as a specific object of analysis. Just the same, the preceding assertion is now commonplace (often supported by a reference to Lippman and Rumelt 1982), appearing in everything from foundational scholarly contributions (e.g., Barney 1991, p. 107; Peteraf 1993, p. 182) to MBA textbooks (e.g., Besanko et al. 1996, p. 552; Collis and Montgomery 1998, p. 34; Grant 2002, p. 238). The point of view taken here is that when "causal ambiguity" is as broadly defined as "the state in which managers do not know how their actions map to consequences," the statement "managers experience causal ambiguity" is indistinguishable from "managers don't know what they're doing," in which case a bias toward plain language should favor the latter.[2] This paper is motivated by the interesting possibility, as initially raised by Lippman and Rumelt (1982), that a *particular type* of confusion can arise in the context of competitive imitation that is both "causal" and "ambiguous" in a precise sense of both words.

To explore this possibility, I create a model in which firms are conceptualized as a collection of activity centers, with differences in firm performance arising from differences in activities undertaken. This is consistent with Porter (1996) and the previously cited work on combinatorial complexity. Here, however, managers cannot choose the entirety of firm-wide activities with deterministic precision but, instead, can only foster influence relations between subordinates who themselves decide which activities get done. Different influence networks generate different probability distributions over activities and, hence, different levels of expected profit. The potential imitator knows neither the stochastic implications of influence structure nor the actual influence relations chosen by an industry exemplar. It does, however, observe the activities of the exemplary firm and, from these observations, may be able to piece together the network of influence relations underlying its successful performance.

Where does causal ambiguity enter the analysis? First, when the behavior of activity centers is consistent with the overarching network of influence relations (in a very natural way), the firm's operations can be represented as a stochastic causal system.[3] As a result, a potential imitator can attempt to learn the influence structure of a rival by applying standard techniques of causal inference to that rival's observed activities. Second, because the imitator begins with subjective beliefs about the stochastic implications of influence relations and about the true structure underlying the superior performance of its target—i.e., beliefs over probabilities—ambiguity is also at play;[4] thus am I able to provide precise definitions of both *subjective causal ambiguity*, pertaining to the imitator's subjective beliefs, and *intrinsic causal ambiguity*, a measure of an influence network's inherent resistance to causal inference.

I demonstrate that the intrinsic ambiguity of a firm's internal influence network presents a long-run upper bound on the subjective ambiguity experienced by its challengers. When financial risk is factored into the analysis, I show that this suggests that imitation is less likely in industries characterized by higher levels of intrinsic ambiguity. Because the relationship between causal ambiguity and the density of influence relations does not increase monotonically, some causally ambiguous strategies may be complex but others deceptively simple. This last point suggests that—contrary to the conventional wisdom—the strategies of small, younger firms may be inherently more difficult to imitate than those of their of larger, more established counterparts. It also implies a normative conclusion that, sometimes, simpler strategies may be better than complex ones. Because the analysis focuses on a firm's observable activities, my propositions are particularly amenable to direct empirical refutation. Finally, I demonstrate that combinatorial complexity and causal ambiguity are, indeed, distinct barriers to imitation. This leads to the nice implication that, to the extent combinatorial complexity and causal ambiguity are both present in the competitive environment, explorative and absorptive learning

[2] The connection between causal ambiguity and imitation has seen some informal refinement (e.g., King and Zeithaml 2001, Mosakowski 1997, Reed and DeFillippi 1990).

[3] These are sometimes referred to as "Bayes nets." Basic introductions to these systems include Pearl (1988, 2000), Edwards (1995), Jensen (1996, 2001), and Korb and Nicholson (2004). More advanced coverage is found in Cowell et al. (1999) and Spirtes et al. (2000).

[4] That is, in the technical sense of Einhorn and Hogarth (1986).

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

are complementary strategic activities, not substitutes. From a positive standpoint, we should expect firms to seek performance enhancement via learning programs of both types.

In the next section, I use some simple examples to illustrate the key concepts. The setup, key assumptions, and essential formal objects are presented in §3. Section 4 introduces a result on the observational equivalence of causal systems from the causal inference literature. Section 5 specifies the degree of managerial rationality assumed in the model. The formal definition of causal ambiguity is presented in §6. The main results are in §7 and §8. Links to the *NK* literature are discussed in §9, and concluding thoughts are provided in §10. A glossary and proofs to the propositions can be found in the online supplement (provided in the e-companion).[5]

2. "Causal Ambiguity" vs. Other Confusions

Consider a simple version of the setup in Lenox et al. (2006). A potential entrant faces an incumbent in a homogeneous-goods market in which price is set according to Cournot competition with linear market demand $p(q^I, q^E) = \alpha - \beta(q^I + q^E)$, where p is price and superscripts I and E indicate the incumbent and entrant, respectively. E's profit function is

$$\pi^E(q^I, q^E, a) \equiv (p(q^I, q^E) - c^E(a))q^E - \kappa, \qquad (1)$$

where κ is a fixed operating cost, and c^E is a constant marginal cost (on q^E) that depends upon an activity profile a. Suppose, the firm has three cost-relevant activities, each of which can be in one of two states. This is represented as $a = (a_1 a_2 a_3)$ with $a_i \in \{\mathbf{0}, \mathbf{1}\}$; i.e., a is a 3-tuple of zeros and ones. For example, if the challenger is an airline, then activities might include the heterogeneity of aircraft types employed, choice of routes, the frequency of service along routes, the specific of frequent flyer programs, passenger boarding procedures, crew decision rights, and so on.[6]

There are $8 = 2^3$ possible joint activity outcomes. Assume the activity–cost relationship is as shown in Table 1. To distinguish between causal ambiguity and combinatorial complexity, this cost function is designed to exhibit *maximal* combinatorial complexity (i.e., an *NK* fitness landscape with $N = 3$, $K = 2$).[7] The cost-minimizing activity configuration is $a = (\mathbf{111})$. Suppose that (i) the challenger can observe the incumbent's activities, (ii) managers can simply set all activities to their desired values, and (iii) the market can support two efficient firms. Then, E enters, implements $a = (\mathbf{111})$ by managerial fiat and immediately succeeds in imitating I.

Table 1 Costs Arising from Specific Activity Outcomes

a_1	a_2	a_3	c^E	a_1	a_2	a_3	c^E
1	**1**	**1**	29	**0**	**1**	**1**	74
1	**1**	**0**	86	**0**	**1**	**0**	63
1	**0**	**1**	43	**0**	**0**	**1**	40
1	**0**	**0**	32	**0**	**0**	**0**	52

Now, consider the ways in which an entrant might misunderstand the situation. It may not know the forms taken by p, π^E, and c^E or the specific values of α and β; whether there are other entrants lurking in the shadows; the actions available to the incumbent; whether the game is Cournot versus Stackleburg; or whether the incumbent irrationally punishes entrants by flooding the market with product. Each of these problems are well studied in game theory (e.g., games of imperfect information, of incomplete information, subjective games, etc.).

In Lenox et al. (2006), the relationship between activities and cost is not known. Firms attempt to learn this relationship via exploration. To see the problem presented by combinatorial complexity, assume E enters and, for whatever reason, chooses $a = \mathbf{000}$ with $c^E = 52$. Suppose E searches "locally" by experimenting with one activity each period. Among the activity profiles in the one-flip neighborhood of **000** (i.e., profiles **100**, **010**, and **001**), E eventually discovers the best, **100**, and improves performance to $c^E = 32$. Searching in the local neighborhood of **100** yields no gains. Hence, in this case, exploration ceases even though the global optimum is not found. Combinatorial complexity acts as a barrier to local exploration.

Any of these confusions could, loosely, be termed "causal ambiguity" in the broadest sense that the entrant does not know the exact consequences of its actions. My goal here is to create a setting in which "causal ambiguity" describes incomprehension with respect to a distinct type of knowledge, one that involves both causality and ambiguity.

What is causal knowledge and why does it matter? One situation in which causal knowledge plays an important role is when agents must choose among interventions designed to shift outcomes in some favored direction. For example, suppose superior performing firms tend to have "high-powered" sales incentive systems coupled with "aggressive" sales cultures. Does the incentive system drive the culture, or is it the other way around? Confusion as to the correct answer might lead a firm to implement

[5] An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

[6] See, e.g., Porter's (1996) in-depth discussion of Southwest Airlines.

[7] Constructed in the standard way, e.g., as in Ghemawat and Levinthal (2000, p. 14).

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

high-powered incentives and let the sales force adapt when it should have hired an aggressive sales force and then adapted the incentive system to them.

A second situation in which causal knowledge is an issue is when influence relations are not an exogenous feature of the environment but, instead, an object of choice. This is the issue studied here. Both Rivkin (2000) and Ghemawat and Levinthal (2000) suggest that, within the firm, influence relations may be malleable. That is, some strategic decisions may actually involve the creation of a causal system. For example, managerial policies may induce an internal network of influence relations across activity centers. Presumably, different networks generate different outcomes—hence, inducing an effective influence network may be a key strategic objective. If so, learning and imitating an exemplar's influence structure may be an important step toward improved performance. If the firm knows where and in what direction the influence relations flow, it is well on its way toward successful imitation. Unfortunately, although the exemplar's operating activities may be observed in great detail, the influence relations driving them may well remain hidden and resistant to discovery, thereby creating a capabilities-based barrier to imitation.

How do managers structure influence relationships between activity centers? At the more tangible end of the spectrum is the organization of work in process between factory units. Organization of this kind is often imposed explicitly, in great detail, and vigorously monitored and enforced. For example, in semiconductor production, manufacturing proceeds through a sequence of activities—planarization, cleaning, etching, diffusion, chemical coating, lithography, oxidation, implantation, sputtering, grinding, polishing, vapor deposition, testing, and spinning—each typically conducted within a specific work unit.

More generally, activity centers (department, shop team, task force, etc.) are influenced by their formal objectives, the individual preferences of their constituent employees (which may be at odds with their formal objectives), the employees' skills and knowledge, the firm's incentive and compensation schemes, and so on. It seems uncontroversial to assert that, in the real world, managers do not have the resources or knowledge required to force each and every activity center to produce a desired outcome every period. Rather, senior managers must rely on a set of fairly blunt organizational instruments to push efforts roughly in the desired strategic direction.[8]

[8] This is in the spirit of Ghemawat and Levinthal (2000, p. 2), who say, "Discussions of cross-sectional linkages often presume that a coherent system of policy choices is arrived at by some process of a priori theorizing . . . . A more plausible characterization is that a firm makes a few choices about how it will compete and these choices, in turn, influence subsequent decisions."

Table 2 **Frequency of Incumbent Operating Outcomes**

a_1	a_2	a_3	Freq. (%)	a_1	a_2	a_3	Freq. (%)
1	**1**	**1**	72.9	**0**	**1**	**1**	0.9
1	**1**	**0**	8.1	**0**	**1**	**0**	8.1
1	**0**	**1**	0.9	**0**	**0**	**1**	0.9
1	**0**	**0**	8.1	**0**	**0**	**0**	0.1

Returning to the example, suppose that managers are limited to issuing activity center targets, creating broad incentives for their achievement, and then hoping for the best. According to Table 1, they will set the target of activity center i to $a_i = 1$. Assume that using the best available incentive schemes, managers induce each of the three activity centers to hit their targets 90% of the time. Then, $\Pr(\mathbf{111}) = 0.9^3 = 73\%$ and, overall, expected cost is 39. Implicit in these calculations is the assumption that the activity centers work *independently* to achieve their respective targets.

Suppose managers can also impose conditional targets. For example, assume I continues to set independent targets for centers 1 and 2 at $a_i = \mathbf{1}$ but now creates the following set of targets for center 3 that are contingent on the activities of units 1 and 2.

$(a_1 a_2)$	a_3 target
(00)	**1**
(01)	**0**
(10)	**0**
(11)	**1**

Without changing any center's ability to hit targets (90% success rate), Table 2 summarizes the frequency of joint activities that is, over time, generated by this organizational strategy. As a result, expected cost performance is improved by 5% over the scheme with independent targets. Qualitatively, the influence relationships created by the new policy are neatly summarized by

$$a_1 \rightarrow a_3 \leftarrow a_2, \tag{2}$$

meaning that activities of area 3 are influenced by the activities of its two independently operating counterparts. When entrant E knows neither (2) nor the performance implications summarized in Table 2, it faces a problem of causal ambiguity.

Suppose I is considered a shining example of superior strategic design and, as such, has had its operations examined in intimate detail—in academic papers, business case studies, and the popular press. Assume that I's operating history is public but that the organizational structure underlying that performance is hidden to outsiders. E, having read all the studies, is aware of the information in Table 2.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

I call these frequencies the *empirical distribution on activities induced by* (2) and denote it by μ.

Assume E does not know the cost landscape, I's organizational strategy (2), nor the rate at which departmental targets are attained. Assume E has access to all the same productive resources available to I (and, hence, could instantly imitate if it knew which influence relationships to establish between activity centers). Does Table 2 provide sufficient information to imitate? The answer, as it turns out, is yes. Moreover, the path to enlightenment, in this case, is provided by what are now standard techniques of causal inference.

One hypothesis that might be entertained by E is that I's activity centers operate independently. If so, μ is fully described by three influence parameters, θ_1, θ_2, and θ_3, each corresponding to a "local" probability associated with the independent operating structure; i.e., $\mu(a_i = \mathbf{1}) = \theta_i$. Then, according to Table 2,

$$\begin{aligned}
\mu(\mathbf{111}) &= \theta_1\theta_2\theta_3 &&= 0.729,\\
\mu(\mathbf{110}) &= \theta_1\theta_2(1-\theta_3) &&= 0.081,\\
\mu(\mathbf{101}) &= \theta_1(1-\theta_2)\theta_3 &&= 0.009.\\
&\vdots \qquad\qquad \vdots && \qquad \vdots
\end{aligned}$$

A moment's reflection should convince E (and us) that these equalities cannot be consistently satisfied by any choice of parameter values. Hence, even though E is ignorant of I's hidden operating structure, it should at least rule out "establish independent operations" as its imitative goal.

Alternatively, suppose E hypothesizes that I has adopted (2). Then, μ is fully described by the set of local influence parameters $\{\theta_1, \theta_2, \theta_{3|\mathbf{00}}, \theta_{3|\mathbf{10}}, \theta_{3|\mathbf{01}}, \theta_{3|\mathbf{11}}\}$ where, e.g., $\theta_{3|\mathbf{00}}$ is the likelihood that $a_3 = \mathbf{1}$ when $a_1 = a_2 = \mathbf{0}$. For these to be consistent with Table 2, they must satisfy

$$\begin{aligned}
\mu(\mathbf{111}) &= \theta_1\theta_2\theta_{3|\mathbf{11}} &&= 0.729,\\
\mu(\mathbf{110}) &= \theta_1\theta_2(1-\theta_{3|\mathbf{11}}) &&= 0.081,\\
\mu(\mathbf{101}) &= \theta_1(1-\theta_2)\theta_{3|\mathbf{10}} &&= 0.009.\\
&\vdots \qquad\qquad \vdots && \qquad \vdots
\end{aligned}$$

As we know (from the construction of Table 2), these parameters do, indeed, exist. Thus, the entrant cannot rule out (2).

By applying this procedure to the remaining possibilities, the *only* influence network that survives is the true one. This is so in spite of the combinatorial complexity of the cost function. Indeed, combinatorial complexity does not enter the preceding analysis in any way. By carefully observing the operations of I, E can infer the efficient structure and imitate I immediately upon entry (without exploration). As we will see, the transparency of this case happens to be very special. More generally, the optimal network of organizational influence relations is causally ambiguous. This ambiguity can foil inferences of the type illustrated above—a result that does not rely on any assumption of short-run fixity of resources or other exogenously imposed technological constraint (i.e., one that is strictly knowledge based). Characterizing this barrier is the task to which I now turn.

3. The Model

To start, assume that a lone incumbent, I, faces a single potential entrant, E. Because I wish to isolate the effects of causal ambiguity from differences in firm capabilities, assume that both firms have access to identical technologies and resource portfolios. In particular, both firms consist of $2 \le r < \infty$ *activity centers*. Firms also share a common discount factor δ.

Competition is dynamic. The timing of events within each period is as follows. E decides whether to enter and, if so, how to organize its influence relations. I always enters and implements the superior organization (as described in the next section). Activities are simultaneously generated for I and, if it enters, E via a stochastic process that is induced by the firms' respective influence structures. These activities determine the marginal costs under which the firms compete Cournot-style to receive their payoffs.

3.1. Activities and Performance Outcomes

An *activity profile* for firm i in period t, denoted a_t^i, is an r-vector of zeros and ones; e.g., the kth component $a_{k,t}^i \in \{0, 1\}$ so that, e.g., $a_t^i = (100\ldots010)$. Broadly, "activities" are externally observed resource state variables indicating the quality, quantities, or prices of productive inputs, inventory levels, plant locations, composition of workforce skills, production processes utilized, and so on. Define A to be the set of all possible activity profiles including a null profile, $a_\varnothing$, to go with the no-entry case.[9]

Let $\pi^i(a_t^I, a_t^E)$ be the period-t profit to firm i when the firms' activity profiles are a_t^I and a_t^E. Assume that (i) there is an efficient activity profile a^{best} such that $\pi^E(a^{\text{best}}, a^{\text{best}}) > 0$, and (ii) there is a profit-minimizing activity profile $a^{\text{worst}} \neq a_\varnothing$ such that $\pi^E(a_t^I, a^{\text{worst}}) = -\kappa$, regardless of the value of a_t^I. Normalize the payoff to staying out of the industry to zero: $\pi^i(a_t^I, a_t^E) = 0$ if $a_t^i = a_\varnothing$ (i.e., not entering implies zero economic profit).[10] Note that the Cournot setup in (1), with NK-tuned costs, is a special case of these assumptions.

[9] $A \equiv \{0, 1\}^r \cup a_\varnothing$. Unless otherwise indicated, all sets are finite.

[10] It should be mentioned that the following results hold in much more general settings, including those with multiple entrants, large activity domains ($a_{k,t}^i \in \{0, \ldots, k\}$, $k < \infty$), and environments in which individual profits depend directly upon a_t (allowing, e.g., activity-contingent product differentiation).

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

3.2. Policy Choices

In this model, the influence network between activity centers is an object of managerial choice (i.e., through organizational policy). I refer to these networks as firm *operating structures* and, formally, depict them with directed, acyclic graphs in which the nodes correspond to the firm's individual activity centers and the edges correspond to the direct influence relationships between them. For example, Intel's arrangement of the 10 semiconductor activities mentioned previously, is described by the graph

$$a^{\text{Intel}}_{1,t} \to \cdots \to a^{\text{Intel}}_{10,t},$$

e.g., the output of *vapor deposition* directly influences *testing results*, $a^{\text{Intel}}_{8,t} \to a^{\text{Intel}}_{9,t}$, whereas *planarization* affects *testing* as well, but *indirectly* through its influence on the outcomes of intermediating processes.[11]

As an object of analysis, operating structure is intended to capture the *actual* influence relationships that drive outcomes. Often, these do not correspond to any official organizational chart but, instead, are tacit and, hence, invisible to outside observers (such relationships are commonly referred to as the "informal organization"). For example, Moody (1995), chronicles a year spent at Microsoft shadowing the Encarta design and development team. As Moody (1995) points out, managers often imposed broad structure on the informal organization in the sense I have in mind here; referring to one senior manager, Moody says (p. 217), "His direct interventions in team disputes invariably were in support of Bjerke—an endorsement, it seemed to me, of the decisions she was making," where Bjerke represented the Design component of this team (the other departments included Development and Marketing). This observation is consistent with

$$\begin{array}{ccc} a_{\text{Design}} & \longrightarrow & a_{\text{Development}} \\ & \searrow & \\ & a_{\text{Marketing}} & \end{array}$$

Any operating structure is permitted, provided it is free of influence loops (acyclic).[12] Structures need not be fully connected nor are they required to be sensible. For example, Intel is allowed to try

$$a^{\text{Intel}}_{10,t} \to a^{\text{Intel}}_{9,t} \to \cdots \to a^{\text{Intel}}_{1,t},$$

presumably with disastrous results.

Index the various operating structures that can be arranged using the r activity centers by $1, \ldots, m$. Robinson (1977) demonstrates that there are

$$m(r) = \sum_{k=1}^{r} (-1)^{k+1} \binom{r}{k} 2^{k(r-k)} m(r-k), \tag{3}$$

directed acyclic graphs that can be constructed from r nodes (where $m(0) \equiv 1$). By (3), $m(1) = 1$, $m(2) = 3$, $m(3) = 25$, $m(4) = 543$, $m(5) = 29{,}281$, and so on.

Let S denote the set of m operating structures available to both firms, with a typical element (structure) denoted S_k and including a *null structure*, $S_\varnothing$. In the initial, entry-organization phase of period t, firm i chooses a structure $S^i_t \in S$ in which the options are *stay out* ($S^i_t = S_\varnothing$) or *enter* using one of the m non-null operating structures (e.g., $S^i_t = S_k$). If $S^i_t = S_\varnothing$, then $a^i_t = a_\varnothing$ is certain.

3.3. Performance Implications of Structure

In the Microsoft case mentioned above, different elements of the *Encarta* team (at the time, code-named *Sendak*) had different agendas (Moody 1995, p. 27): "*Sendak*'s designers and editors would want to pack the encyclopedia with features seen nowhere else... *Sendak*'s developers would want a far less ambitious set of new features and ample time in which to write code for them." In this situation, "...*which element held sway* would largely determine the functionality of the software, the timing of its release and, ultimately, its success in the marketplace" (emphasis added). In this model, the strategic decision facing managers is determining which activity centers "hold sway" over one another. Inevitably, productive activities of all kinds are prone to a certain measure of unpredictability. In addition to independent "noise" at the local level, I assume activity likelihoods vary in systematic ways with choice of influence structure.

To represent these effects, assume that once an operating structure is established, it generates activities according to a stochastic process along the lines presented in the preceding examples. Suppose $S^i_t = S_k$. The *empirical distribution generated by* S_k, denoted μ_k, is a probability distribution on A. Thus, all firms implementing structure S_k experience the same expected operating performance. Rather than keeping track of the local influence parameters associated with S_k (i.e., the θ.s of the introductory example) and then using them to construct μ_k, I simply take μ_k as a primitive and make an assumption that guarantees the existence of parameter values that will generate μ_k in

[11] Narrowly interpreted, the notion that managers can precisely establish stable influence relationships between activity centers is a heady one. More broadly, imagine that managers choose from a menu of monitoring, enforcement, incentive, and compensation policies that, as a result, induce a de facto operating structure—the latter being the object of analysis here.

[12] Clearly, most productive activities undertaken by firms involve feedback loops between resource units. Note that in such cases, loops can often be eliminated via an appropriate choice of time period and action labeling. Also, although simplifications have been made for the purpose of this paper, the theory of causal inference is sufficiently rich to relax this assumption, including allowing for hidden variables and bidirectional influence relations.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

the desired way. This assumption is what makes the activity centers behave as a causal system; it creates a link between structure and activity that makes causal inference possible.

Assumption 1 (Faithfulness). *For all $S_k \in S$ and $i = 1, \ldots, r$,*

1. *a_i is μ_k-conditionally independent of all its nondescendents given the outcomes of its parents in S_k and*
2. *The removal of any edge in S_k causes item 1 to fail for some a_i.*

For example, if μ is faithful to (2), then it is the case that, for all $a \in A$, $\mu(a) = \mu(a_1)\mu(a_2)\mu(a_3 \mid a_1, a_2)$ but *not* $\mu(a) = \mu(a_1)\mu(a_2)\mu(a_3)$.[13] I also assume that the μ_k are positive on A (i.e., managers cannot eliminate undesired action profiles by choice of structure). Let F_k denote the set of empirical distributions that are faithful to S_k (i.e., for $S_k \neq S_\varnothing$).

Let $\overline{\pi}_k^E$ denote the expected profit for E when it chooses entry structure S_k (i.e., before actual activity profiles are generated),

$$\overline{\pi}_k^E \equiv \sum_{a^I \in A} \sum_{a^E \in A} \pi^E(a^I, a^E)\mu_1(a^I)\mu_k(a^E), \tag{4}$$

where $\pi^E(a^I, a^E)$ meets the conditions enumerated in §3.1. Assume that there is a strictly dominant operating structure in the sense of maximizing expected profit regardless of the operating structure of one's competitor. Label this S_1. Fix the incumbent's actions to $S_t^I = S_1$ for all t. Hence, E certainly enters if imitation is guaranteed. I prefers that E stay out. Barring that, I prefers that E choose an inefficient operating structure because I's profit is inversely proportional to E's cost.[14]

4. Observational Indistinguishability Theorem

S_l is said to be *observationally indistinguishable from* S_k if the empirical distribution generated by S_l is also faithful to S_k (i.e., if $\mu_l \in F_k$). To see why this distinction is important, suppose that, over a sufficient period of observation, the challenger develops an arbitrarily accurate assessment of the empirical distribution on the incumbent's activities, μ_1. If $\mu_1 \in F_k$ where $k \neq 1$, the challenger (who does not observe the incumbent's actual operating structure, S_1) cannot tell which of S_1 or S_k is actually generating the results. This is where the opportunity for confusion arises. E might choose to implement S_k, in which case direct experience will eventually reveal that it is not the optimal structure.[15] However, if E is worried that the incorrect choice results in very poor performance, it might choose to skip entry and stick with something safer (i.e., its known outside alternative). Alternatively, as I will show, if $\mu_1 \notin F_k$, then S_k can, over time, be ruled out strictly via passive observation.

Let OI_k denote the set of structures with which S_k is observationally indistinguishable. Under the faithfulness assumption, $S_l \in OI_k$ if and only if $\mu_l \in F_k$. Thus, if S_1 is completely "transparent" in the sense that no other structure is observationally indistinguishable from it ($OI_1 = \{S_1\}$), then it is generally only a matter of time before the challenger properly identifies the efficient organization and enters. As we will see, operating structures are, in general, not so transparent.

Consider, for example, the simple operating structure $a_1 \leftarrow a_2 \rightarrow a_3$. Suppose that the influence parameters of are $\theta_2 = 0.90$, $\theta_{1|0} = 0.10$, $\theta_{1|1} = 0.90$, $\theta_{3|0} = 0.80$, and $\theta_{3|1} = 0.90$. Then, the empirical distribution on activity profiles is

a_1	a_2	a_3	μ	a_1	a_2	a_3	μ
1	1	1	72.9	0	1	1	8.1
1	1	0	8.1	0	1	0	0.9
1	0	1	0.8	0	0	1	7.2
1	0	0	0.2	0	0	0	1.8

(5)

However, this *same* distribution is implied under $a_1 \rightarrow a_2 \rightarrow a_3$ by parameters $\theta_1 = 0.82$, $\theta_{2|0} = 0.50$, $\theta_{2|1} = 0.98$, $\theta_{3|0} = 0.08$, and $\theta_{3|1} = 0.90$. Similarly, $a_1 \leftarrow a_2 \leftarrow a_3$ induces this distribution when $\theta_3 = 0.90$, $\theta_{2|0} = 0.90$, $\theta_{2|1} = 0.90$, $\theta_{1|0} = 0.90$, and $\theta_{1|1} = 0.90$. Thus, an observer of (5), could not tell which of these three operating structures actually generated the data.

It would certainly be useful if the observational indistinguishability class of the incumbent's operating structure could be constructed directly from the features of its influence network (i.e., and not require "brute force" construction by repeated application of Bayes rule to all m possibilities). Fortunately, as it turns out, this is possible. For the following theorem, given a structure S_k, three activities are said to constitute a *local structure identifier* (LSI) if two unlinked activities are organized to influence the third directly, e.g., a structure like $a_1 \rightarrow a_3 \leftarrow a_2$.

Theorem 1 (Verma and Pearl 1990). *Two organizational structures are observationally indistinguishable if*

[13] See Spirtes et al. (2000, p. 13) for additional technical details. Faithfulness turns out to be a "reasonable" assumption in the sense that, under mild regularity conditions, of all possible parameter values ($\theta_{\cdot}$), the set failing the faithfulness condition has Lebesgue measure zero (Meek 1995).

[14] Loss-assuring a^{worst} implies, for all S_k, that there are faithful empirical distributions, $\mu_k \in F_k$, under which $\overline{\pi}_k^E < 0$.

[15] Over time, E learns that its performance is, on average, different from I's ($\mu_k \neq \mu_1$). Because S_1 is assumed to be uniquely efficient, it also discovers $\overline{\pi}_k^E < \overline{\pi}^I$.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

and only if they have the same edges (*regardless of direction*) *and set of LSIs*.

Suppose, for example, that the incumbent's organization is $a_1 \leftarrow a_3 \rightarrow a_2$. Using Theorem 1, we determine, via visual inspection alone,

$$OI_1 = \{a_1 \leftarrow a_3 \rightarrow a_2, a_1 \rightarrow a_3 \rightarrow a_2, a_1 \leftarrow a_3 \leftarrow a_2\}. \quad (6)$$

To see this, first note that a structure is observationally indistinguishable with $a_1 \leftarrow a_3 \rightarrow a_2$ only if it has the same edges: the only possibilities are those shown in (6) plus $a_1 \rightarrow a_3 \leftarrow a_2$. None of the structures in (6) are ruled out because they all have the same set of LSIs (the empty set). However, this is not true of $a_1 \rightarrow a_3 \leftarrow a_2$ because, as configured, this structure constains one LSI (in this case, the graph itself).

Example 1 (Gatekeeper). Alternatively, consider the following structure involving five activity centers: Budgeting, Engineering, Finance, Manufacturing, and Marketing.

$$\begin{array}{ccccc} Bud & & & & Eng \\ & \searrow & & \swarrow & \\ & & Fin & & \\ & \swarrow & & \searrow & \\ Mfg & & & & Mkt \end{array} \quad (7)$$

Here, Financial Analysis serves as a gatekeeper, checking Engineering projects against Budget's projections before forwarding approved projects to Manufacturing and Engineering. By Theorem 1, there are no other structures in the observational indistinguishability class: keeping the edges constant, there is no way to reverse an arrow without breaking an LSI or forming a new one.

This result is important because it demonstrates a general insight on strategic inference with respect to organizational influence: *larger and/or denser networks can make the organization more transparent*. Moreover, Theorem 1 tells us exactly what kinds of relationships serve to increase this transparency.

5. Subjective Rationality

To keep our attention on the causal inference problem, assume that E is aware of all environmental primitives except (i) which structure is the efficient one consistently chosen by I, and (ii) the actual empirical distributions associated with each of the m available operating structures.[16] In this scenario, E is far better informed than most real-world counterparts would be under similar circumstances. For example, it knows what its resources are, how to parameterize market demand, how to organize its activity centers to create the desired influence relationships, and so on. Still, a crucial piece of the strategic puzzle—which operating structure is efficient—is missing.

Assume that, following each period, E observes I's activity profile. E also knows its own entry/organization decisions and activity profile outcomes. Therefore, at the start of period t, E observes a *history* of the form $h_t \equiv (h_0, S_1^E, a_1^I, a_1^E, \ldots, S_{t-1}^E, a_{t-1}^I, a_{t-1}^E)$, where h_0 is the null history.

I adopt the notational convention of using hats to indicate E's assessment with respect to an object. So, whereas $\mu \equiv (\mu_1, \ldots, \mu_m)$ summarizes the true empirical distributions associated with each of the m possible structures, $\hat{\mu}_0 \equiv (\hat{\mu}_{1,0}, \ldots, \hat{\mu}_{m,0})$ summarizes E's initial beliefs about them. Let $\hat{\rho}_0 \equiv (\hat{\rho}_{1,0}, \ldots, \hat{\rho}_{m,0})$ be E's initial assessment about which structure is optimal; e.g., $\hat{\rho}_{k,0}$ is E's period-0 belief that S_k is the true efficient structure (the one employed by I). $\hat{\rho}_0$ is a set of beliefs that is used to weigh other beliefs, thereby adding the ambiguity dimension to the model.[17] Let $\hat{\mu}_k(h_t)$ be E's updated assessment of μ_k given the history h_t. When the history is implied, I write $\hat{\mu}_{k,t}$ (when it is clear from the context, I further simplify notation by dropping the time subscripts on $\hat{\mu}$ and $\hat{\rho}$ altogether). E is savvy to the idea of using causal inference to infer I's operating structure: it knows that, for any S_k, $\mu_k \in F_k$. Finally, assume that E's initial priors are independent, Dirichlet distributed, and result in strictly positive multinomial distributions.[18]

From E's perspective, there are m relevant "states of the world"—one corresponding to each structure in which that structure is the optimal choice ($\hat{\rho}$ weighs these states each period). E may have different assessments conditional upon which state of the world it finds itself.

A *dynamic strategy*, denoted σ, for firm E specifies a (possibly random) entry/organization choice for every history: formally, $\sigma(S_k \mid h_t)$ is the probability that E chooses $S_t^E = S_k$ upon observing the history h_t; σ can encode simple strategies ("stay out forever," "enter in even periods under S_k," etc.) as well as much more sophisticated, outcome-dependent ones ("fix $\varepsilon \in [0, 1]$ and enter with S_k in any period t where $\hat{\rho}_{k,t} \geq 1 - \varepsilon$," "employ NK-style search from period t on," "embark on a subjectively optimal program of Bayesian experimentation," etc.).

Assume E is *subjectively rational*, meaning that (i) beliefs are updated in Bayesian fashion, and

[16] Previous versions of this paper included uncertainty about the general mapping from activity configurations to payoffs and, in the case with multiple entrants, the strategies of rivals. The results were virtually identical at the expense of a massive increase in mathematical complexity.

[17] In this formulation, $\hat{\mu}$ depends upon $\hat{\rho}$. See the online proofs supplement for the technical details.

[18] The Dirichlet assumption is primarily to ensure that E can, indeed, update its beliefs in response to new information. Interested readers are referred to Neapolitan (2004, p. 309).

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

(ii) σ maximizes the subjective expected present value of profits in every period.[19] Given the assumptions, entry from period 1 on is, in fact, the optimal strategy. The only way imitation is forestalled, therefore, is if E's subjective assessments are persistently wrong. If E is allowed to believe anything (e.g., "Martians always strike down imitators with death rays") then finding beliefs that cause imitation to fail is trivial. However, E is not only a subjective optimizer, but also a rational learner—its beliefs are properly updated in response to new information. Thus, it is not obvious that there are *any* initial beliefs that, ultimately, cause imitation to fail.

6. Causal Ambiguity

Given this setup, it is possible to introduce two measures of causal ambiguity, one with respect to the inherent transparency of the incumbent's operating structure and another with respect to a firm's subjective beliefs regarding which structure that is. Given perfect knowledge of μ_1 (I's empirical distribution), there is a limited number of structures that E might confuse with S_1 under the faithfulness assumption. The idea is to relate this number to the subjective beliefs of E over time under Bayesian learning. Presumably, E's beliefs must at least converge (almost surely) to place positive weight only on structures faithful to S_1. Less obvious is whether, under subjective rationality, the opportunity to enter and experiment with structures of its own implies that E must *inevitably* learn the efficient structure.

Definition 1. Given beliefs $\hat{\rho}$, E's *subjective degree of causal ambiguity* is

$$\hat{\lambda}(\hat{\rho}) \equiv -\sum_{k=1}^{m} \hat{\rho}_k \ln(\hat{\rho}_k). \tag{8}$$

This measure ranges from zero to $|\ln(m)|$ (a positive number).[20] It equals $|\ln(m)|$ when the challenger's priors regarding the optimal structure are uninformative (i.e., $\hat{\rho}_1 = \cdots = \hat{\rho}_m = 1/m$) and zero when the challenger is certain that it knows which structure is the optimal one.[21] This definition is useful because it summarizes each firm's uncertainty regarding the efficient structure in a single number. Let m_1 be the number of structures that are observationally indistinguishable from S_1 (i.e., the set cardinality of OI_1).

Definition 2. The *intrinsic ambiguity of* S_1 is $\lambda^* \equiv |\ln(m_1)|$.

Intrinsic ambiguity is equal to the subjective degree of causal ambiguity when managers place equal weight on, and only on, the elements of a structure's observational indistinguishability set. Other than the requirement that E not initially rule out any structure from potentially being the efficient one, $\hat{\rho}$ is fairly unrestricted. Therefore, it is not obvious what, if any, long-run relationship exists between $\hat{\lambda}$ and λ^*. For example, if the subjectively optimal strategy specifies entry and experimentation until the true structure is discovered, $\hat{\lambda}$ must eventually converge to 0.

Suppose, for example, that there are three structures in the observational indistinguishability class of S_1. Then $\lambda^* = \ln(3) = 1.1$. Suppose that $r = 3$ and that E has uninformative initial beliefs (places equal weight on each of the $m = 25$ possible structures that it is the efficient one). Then, $\lambda(\hat{\rho}) = \ln(25) = 3.2$. Alternatively, the structure in Example 1 has $\lambda^* = 0$ even though, in total, there are $m = 29{,}281$ ways to organize these units. Compare this against $a_1 \leftarrow a_3 \rightarrow a_2$, which has $\lambda^* = 1.1$ even though $m = 25$.

7. Result on Long-Run Ambiguity

The questions of interest in this section include the following, assuming E stays out and, hence, gains no first-hand knowledge regarding the structure–performance relationship: To what extent does E learn to predict the I's operating activities? How accurate does E's assessment eventually become regarding the efficient structure? What is the relationship between intrinsic ambiguity and E's long-run subjective ambiguity?

First, how well does E come to predict I's operating behavior? Let $\hat{\mu}_t^I$ denote E's period-t assessment of the incumbent's true empirical distribution. Then, the answer, provided by the next lemma, is that $\hat{\mu}_t^I$ converges to reality with probability 1.[22] Because I's activities are stochastically determined, it is always possible that the actual history observed by E will, by pure chance, happen to mimic data driven by some distribution other than μ_1. What the lemma says is that, over time, large discrepancies between E's beliefs and the truth are highly unlikely. This degree of learning occurs even under a strategy of strictly passive observation (E stays out forever).

Lemma 1. *For all strategies σ and initial beliefs $\hat{\mu}$, E's subjective assessment $\hat{\mu}_t^I$ converges in probability to μ_1. Formally,*

$$\operatorname*{plim}_{t\to\infty} \hat{\mu}_t^I = \mu_1.$$

[19] This particular notion of subjective rationality was introduced by Kalai and Lehrer (1993). Ryall (2003) provides the first application to strategy.

[20] We set $0 \cdot \ln(0) \equiv 0$.

[21] λ is the *entropy* measure of $\hat{\rho}$ (see, e.g., Golan et al. 1996).

[22] The reference distribution in these propositions is always reality (i.e., the distribution over histories implied by the policy choices of I and E and the true empirical distributions associated with each operating structure).

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

Lemma 1 and Theorem 1 imply that E's beliefs regarding I's operating structure are virtually certain to become concentrated with weight 1 on the set of structures in S_1's observational indistinguishability class, OI_1. Thus, because causal structures induce specific, well-defined regularities in the observations they induce, shrewd scrutiny of incumbent conduct cannot but help to reduce the number of structures considered likely candidates for imitation. This important result is stated formally in the following proposition.

Proposition 1. *For all strategies σ and initial beliefs $\hat{\mu}$, E's subjective likelihood that $S_t^I \in OI_1$ converges in probability to* 1:

$$\operatorname*{plim}_{t\to\infty} \sum_{S_t^I \in OI_1} \hat{\rho}_{k,t} = 1.$$

As we now know, more activities permit more linkage choices, which can, in many cases, make the overall structure *more transparent*. More generally, E's subjective ambiguity is limited by the intrinsic ambiguity of I's operating structure as described in Corollary 1, below.

Corollary 1. *For all strategies σ and initial beliefs $\hat{\mu}$, E's subjective degree of causal ambiguity converges in probability to a number bounded by the intrinsic ambiguity of S_1*:

$$\operatorname*{plim}_{t\to\infty} \hat{\lambda}(\hat{\rho}_t) = x \le \lambda^*.$$

Even the challenger that never enters and, hence, never experiments with operations of its own, eventually learns the incumbent's empirical distribution to an arbitrary degree of accuracy. By Theorem 1, this limits beliefs with respect to what the incumbent is actually doing to generate its observed behavior. Of course, under sufficient experimentation, ambiguity may be reduced even further and, perhaps, eliminated altogether. However, even these results are sufficient to refine the *NK* hypothesis.

Corollary 2. *E's subjective degree of causal ambiguity converges in probability to a value that is not monotonic in the combinatorial complexity of the profit function.*

Corollary 2 makes an important point about causal ambiguity in the real world, one that slips through the intuitively appealing reasoning employed in traditional discussions on this topic. Causal ambiguity arises as an issue in strategy because it is thought to be a source of the kind of confusion that prevents managers from imitating the performance of their more successful competitors. As we see in (3), the aggregate number of feasible causal structures is, indeed, exponentially increasing in the number of observed activities. However, the very context of the problem—posited as a challenger attempting to imitate a successful incumbent—already implies that the challenger observes certain aspects of the exemplary firm's behavior. This, in turn, implies the possibility of applying the tools of causal inference to the observed history. Over time, E learns μ_1 and, in turn, the observational equivalence class of structures capable of generating it.

Sometimes problems with a lot of choices are hard to solve and sometimes they are easy—it all depends upon the ruggedness of the landscape. Similarly, if points on a rugged landscape emit data according to a location-specific causal process, then sometimes the area of search is small and sometimes large—it all depends upon the degree of causal ambiguity. As in (7), the lucky case for the entrant is when the incumbent's location can be narrowed down to a single point. Such cases may be rare because the intrinsic ambiguity of most operating structures is greater than zero. Of course, once causal inference is taken as far as it will go, E may very well wish to enter and apply active learning procedures to assess the remaining options. The fact that $\lambda^* < |\ln(m)|$ implies that learning by observing *always* results in a reduction in the search space.

8. Result on Sustained Advantage

Because E controls the same technology as I, any sustainable advantage for the incumbent is "capability based" in the sense of Saloner et al. (2001, pp. 41–55). Therefore, I is said to sustain a *strong capabilities-based advantage* if E never enters. This is a very strong form of imitative failure. In this case, the incumbent's advantage with respect to the potential entrant is sufficient to guarantee it monopoly profits. Alternatively, a *weak* advantage is one in which E never imitates. I would enjoy a weak advantage if E attempted imitation, found a profitable but suboptimal structure and decided the risks of further organizational innovation offset the perceived benefits. A strong advantage implies a weak advantage but not conversely (hence, the terminology).

To decide its course of action, E must conduct a subjectively rational risk analysis. That is, E must convert its beliefs regarding its various organizational options—whatever their degree of ambiguity—into an appropriate profit assessment. Ambiguity by itself is never sufficient to deter entry or imitation. For example, an entrant may have strong expectations that entry is profitable under any of the firm's feasible range of activity profiles (i.e., the entrant's priors place high probabilities on good outcomes under every choice of structure). Even if this assessment is overly optimistic, entry occurs and, at best, only a weak advantage obtains. However, if entry is perceived to be sufficiently risky, it may be deterred altogether (with causal ambiguity playing a key supporting role). To capture these considerations, I now introduce the following risk measure.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

Definition 3. The *intrinsic risk* of S_1 is

$$\xi \equiv -\left(\frac{1}{m_1}\overline{\pi}_1^E - \frac{m_1-1}{m_1}\kappa\right). \tag{9}$$

It is important to note that this measure does not depend upon E's subjective beliefs—it is computed from the objective primitives of the model. To interpret ξ, consider a situation in which sufficient time has passed that E's beliefs regarding μ_1 are arbitrarily accurate (as Lemma 1 guarantees). Suppose that no entry has occurred up to this point. E knows that there are m_1 structures in OI_1 and that one of these structures is the one generating μ_1. What E does not know is which empirical distributions go with which structures. In this case, a lower bound on the worst possible payoff should E choose unwisely is $-\kappa$.[23] ξ is the Bayes risk (see DeGroot 1970, pp. 121–123) of an extremely pessimistic challenger who knows the true empirical distribution is μ_1 but has uninformative beliefs with respect to which structure in OI_1 is the efficient one.

Proposition 2. *Assume that E knows $\mu^I = \mu_1$. If $\xi < 0$, then I does not enjoy a strong capabilities-based advantage nor, with probability* 1, *does it enjoy a weak capabilities-based advantage.*

If $\xi < 0$, then entry happens immediately because even the most pessimistic beliefs cannot deter it. Indeed, the situation is even worse: E not only enters but persists in experimenting until imitation succeeds. Although I may enjoy a weak advantage for some time, it is highly unlikely to last forever (i.e., there are sequences of random events under which E never imitates, but they occur with probability 0). Eventually, E figures out that its current structure is not performing to expectation and tries something different (because $\xi < 0$, E does not exit). Notice the connection to intrinsic ambiguity. When intrinsic causal ambiguity is zero, $m_1 = 1$ and $\xi < 0$. This implies the following corollary.

Corollary 3. *Assume that E knows $\mu^I = \mu_1$. If $\lambda^* = 0$, then I does not enjoy a strong capabilities-based advantage nor, with probability* 1, *does it enjoy a weak capabilities-based advantage.*

The point of Corollary 3 is that, regardless of the financial risk, E will always enter and, with virtual certainty, successfully imitate the incumbent when the operating structure of the latter is intrinsically transparent. Essentially, this structural transparency eliminates financial risk by making the likelihood of successful imitation enormous relative to the likelihood of failure.

It is very important to note that, when $\xi \geq 0$, it is not hard to construct subjective beliefs that support E staying out with probability 1. This holds even though, in this setup, the direct cost of entry/imitation is zero. Interestingly, subjectively optimal experimentation—as *must* arise under the assumption that E is subjectively rational—is not sufficient to assure objectively optimal behavior. Corollary 3 confirms the conventional wisdom in strategy insofar as causal ambiguity is a necessary condition for imitation to fail. However, it is not sufficient: if the financial benefit of successful imitation is too strong relative to the cost of failure, challengers enter and doggedly experiment with organizational structures until they get it right. To gain some intuition into these effects and their relationship to ξ, let us turn to a final example.

Example 2. Consider a situation in which there are two activity centers ($r = 2$). Let $S_1 = a_1 \rightarrow a_2$, $S_2 = a_1 \leftarrow a_2$ and $S_3 =$ (independent operations). By the indexing convention, $S^I = S_1$. For the sake of simplicity, limit the number of periods to two. Should E imitate successfully, it enjoys an expected payoff in each period of $\overline{\pi}_1^E > 0$. E does not know that $S^I = S_1$, but let us suppose it is sufficiently informed to have a very accurate assessment of I's operating behavior; i.e., $\hat{\mu}^I = \mu_1$.

Let $\hat{\rho}_1, \hat{\rho}_2$, and $\hat{\rho}_3$ be E's initial priors on $S^I = S_1$, $S^I = S_2$, and $S^I = S_3$. Because E knows that $\mu^I = \mu_1$ is generated by a causal system, it knows at least that $S^I \in OI_1 = \{S_1, S_2\}$. Therefore, $\hat{\rho}_3 = 0$ and $\hat{\rho}_1 + \hat{\rho}_2 = 1$. In addition to not knowing which structure is the optimal way to organize its two activity centers, it also does not know the empirical distribution associated with the wrong choice. Assume that E is very pessimistic: it believes the wrong choice of structure generates the bad activity profile a^{worst} with certainty, resulting in an expected loss equal to fixed operating costs κ.[24]

Now, from E's perspective, if it enters and chooses wisely, it enjoys expected profit of $\hat{\pi}^E_{\text{good}}$, where $\hat{\pi}^E_{\text{good}} = \overline{\pi}_1^E = 1$. However, if it chooses unwisely, it suffers an expected loss of $\hat{\pi}^E_{\text{bad}} = -\kappa$. To simplify things even further, assume that exploration is incredibly effective; specifically, if E enters in period 1, then it learns beyond a doubt whether its organization is the efficient one. Of course, if E stays out, it learns no additional information.

E has three logical choices: (i) stay out, (ii) enter under S_1, and (iii) enter under S_2. If staying out is optimal in period 1, given the fact that E learns nothing new, it is also optimal in period 2. The net present value of the stay out strategy is, therefore, 0. If E

[23] For each structure S_k and all $\varepsilon > 0$, there exists an empirical distribution $\mu_k \in F_k$ such that $\overline{\pi}_k^E = \varepsilon - \kappa$.

[24] More properly, E believes a^{worst} occurs with probability $1 - \varepsilon$ for $\varepsilon > 0$ arbitrarily small, a technical detail I ignore for the purpose of the example.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

enters under S_1, then, with probability $\hat{\rho}_1$ it earns period 1 profits of $\hat{\pi}^E_{\text{good}}$, and with probability $(1-\hat{\rho}_1)$ a loss of $\hat{\pi}^E_{\text{bad}}$. However, by entering in period 1, it learns the correct structure and, therefore, is assured a payoff of $\hat{\pi}^E_{\text{good}}$ in period 2. Recalling that the discount rate is δ, the subjective expected present value of this plan is

$$\begin{aligned} V_1 &\equiv \hat{\rho}_1 \hat{\pi}^E_{\text{good}} + (1-\hat{\rho}_1)\hat{\pi}^E_{\text{bad}} + \delta \hat{\pi}^E_{\text{good}} \\ &= (\hat{\rho}_1 + \delta)\overline{\pi}^E_1 - (1-\hat{\rho}_1)\kappa. \end{aligned}$$

Similarly, E's assessment of the present value of entering under S_2 is

$$V_2 \equiv (\hat{\rho}_2 + \delta)\overline{\pi}^E_1 - (1-\hat{\rho}_2)\kappa, \quad \text{where } \hat{\rho}_2 = (1-\hat{\rho}_1).$$

Because E is subjectively rational, if it does enter, it chooses the structure corresponding to the larger of V_1 or V_2. This is entirely determined by the larger of $\hat{\rho}_1$ or $\hat{\rho}_2$. Let $V_k = \max\{V_1, V_2\}$. Then, because it can stay out and be assured a payoff of 0, E does not enter if $V_k < 0$, that is

$$(\hat{\rho}_k + \delta)\overline{\pi}^E_1 - (1-\hat{\rho}_k)\kappa < 0, \tag{10}$$

where, because S_k is the value-maximizing entry choice, $\hat{\rho}_k \geq \frac{1}{2}$. Rearranging terms,

$$\frac{1}{2} \leq \hat{\rho}_k < \frac{\kappa - \delta\overline{\pi}^E_1}{\kappa + \overline{\pi}^E_1}. \tag{11}$$

This example highlights several insights that carry through to the more general case. First, condition (11) and, hence, the possibility of failed imitation only arises as a result of causal ambiguity. If E knows that $S^I = S_1$ (i.e., $\hat{\rho}_1 = 1$), it always enters and earns $(1+\delta)\overline{\pi}^E_1$. Second, entry offers the opportunity to gain additional knowledge via direct experience. Here, all of the learning happens in one period. This is unrealistically fast, but is consistent with what happens over longer periods (with high probability). Thus, the pessimistic E must trade off the benefit from learning (here, getting $\overline{\pi}^E_1$ for certain in period 2) with the downside of implementing the wrong structure in period 1. From condition (11), we see that E is less likely to enter the more impatient it is (low δ) and the lower the relative benefit to learning (i.e., the size of $\overline{\pi}^E_1$ relative to κ). In particular, when $\kappa - \delta\overline{\pi}^E_1 < \frac{1}{2}$, the payoff to experimentation is sufficiently large that E always enters.

Tying the example back to our main result, note that $m_1 = 2$ so that

$$\xi = -\left(\tfrac{1}{2}\overline{\pi}^E_1 - \tfrac{1}{2}\kappa\right).$$

Suppose $\xi < 0$. Then, $\overline{\pi}^E_1 - \kappa > 0$. But, from (10) and the fact that $\hat{\rho}_k \geq \frac{1}{2}$, this implies that E's subjectively best choice of entry structure (whichever it is) results in a strictly positive expected payoff in period 1. Therefore, there is no trade-off to make—entry is the strictly dominant, subjectively optimal decision! Keep in mind, we constructed E's beliefs to be maximally pessimistic given its knowledge of I's operating performance. Thus, if E enters under these beliefs, it enters under *any* beliefs (consistent with $\hat{\mu}^I = \mu_1$). Moreover, even in the more general case, E never exits the market. It continues to learn and adopt subjectively optimal structures until (with probability 1) it succeeds in imitating.

Let me conclude this section with the following observation. Because ξ is constructed from objective primitives, Proposition 2 is, in theory, empirically refutable. In the simplest setting, this requires estimating μ_1 and $\overline{\pi}^E_1$ from incumbent operating data. The estimate of μ_1 then implies OI_1 and, hence, m_1. Finally, κ would be estimated as the expected profit in the worst-case organizational scenario. In the real world, the analysis is complicated by hidden causal relationships, multiple firms, etc. However, analytic techniques do exist for estimating causal structures from the data they generate under these complications. Refuting the corollary is a somewhat simpler affair, requiring "only" the estimation of μ_1 and comparing the resulting λ^* to some measure of imitative success within the industry of study.

9. Causal Ambiguity vs. Combinatorial Complexity

As illustrated by the introductory example, my model can be specified to extend Lenox et al. (2006) in a very transparent way, thereby facilitating comparisons with the growing number of strategy applications that utilize the *NK* formalism. Specifically, with profits determined by the equilibrium in a Cournot game with activity-determined costs that are *NK* tuned, the preceding results in no way depend upon any choices for N and K. This demonstrates that causal ambiguity is, indeed, a distinct barrier to imitation. Still, comparisons must be made with care. Here, unlike Lenox et al. (2006), activities in my model are (quite purposefully) not managerial choice variables. Thus, on one hand, I have shown a set of circumstances under which managers facing an *NK*-complex cost landscape can avoid the correspondent search problem by applying causal inference to the operations of the incumbent.[25] On the other hand, as the astute reader might (rightly) point out, this result is achieved by shifting the object of managerial choice from action profiles to operating structures, thereby

[25] When $\lambda^* = 0$, E need not know anything about the mapping from action profiles to costs (or even profits) to imitate successfully.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

causing a mismatch on a very critical dimension. As Rivkin (2000) is careful to emphasize, the issue is the complexity of the decision problem; i.e., the mapping from decisions/policies/choices to payoff outcomes.[26] Thus, because the only policy variable here is choice of operating structure, one might suspect that the original, complex decision problem was simply replaced with a relatively simple one.

To see that the essential message of this paper with respect to imitation withstands this observation, it suffices to consider an instance of the model in which the choice of operating structure is, itself, *NK* complex. To make things concrete, return to the three-activity case. By (3), there are 25 possible operating structures that meet the acyclicality requirement (not counting the stay-out option). Arbitrarily index these 1 to 25 and assign each their number in five-digit binary (e.g., structure #1 = 00001, #9 = 01001); the optimal structure need no longer be #1. As we know, assuming the incumbent chooses the optimal structure (whichever one it is), each five-digit string corresponds to an expected payoff; e.g., if the optimal structure happens to be #2, the entrant gets an expected payoff of $\overline{\pi}^E(S_{00010}, S_{01001})$ when it picks structure #9. It should be easy to see that by manipulating the demand parameters, cost function, and activity probabilities, we can tune $\overline{\pi}^E$ to any level of *NK* complexity.[27]

My model is sufficiently general to allow E to adopt a strategy σ in which it enters and pursues an exploration strategy using an *NK*-style hill-climbing algorithm. Indeed, the model *requires* that E do exactly this whenever subjective rationality demands it. At the same time, because σ must be optimal with respect to E's updated beliefs in every period, the implications of causal inference must also be respected. That is, to the extent that causal inference rules out certain structures, these must be removed from the set upon which σ searches. Therefore, my results complement the *NK* studies by saying something about the likely "area of the landscape" upon which firms in a particular industry search (as well as how that area changes over time).

Viewed in this way, the theory implies that explorative and absorptive learning strategies should not be viewed as substitutes but as complements, each defeating a different type of learning barrier. My results demonstrate that passive learning (e.g., accumulating competitive intelligence) inevitably shrinks the search space upon which active learning operates. Conversely, active learning (e.g., attempts at de novo innovation via experimentation) create new information that cannot but help to refine a firm's understanding of how the world works, thereby improving its ability to decipher the behavior of an industry's superior performers.

From a positive point of view, subjective rationality implies both of these approaches are constantly being weighed, with efforts on each typically being applied contemporaneously. Thinking of firms as employing either an active or passive imitation strategy is too narrow. When intrinsic ambiguity is low, more emphasis may be placed on the latter (only in the extreme case of zero ambiguity is exploration entirely uncalled for). When ambiguity is high, less is learned by observing, the benefits to exploration are greater and, as a result, we should observe more of it. This also implies some caution in inferring firm strategies from *outcomes*: innovative search on a landscape narrowed by causal induction can produce outcomes that—to the outside observer—look either more like imitation or more like innovation.

10. Conclusions

On one hand, the preceding results confirm the conjecture that causal ambiguity may well play an important role in developing a capabilities-based advantage. On the other, its mere existence is not sufficient to ensure it. Given enough experimentation, an entrant eventually discovers the optimal operating policy. More patient firms place greater weight on the benefits of experimentation and, hence, are more likely to adopt exploration strategies. Firms with a fairly high level of confidence (lower subjective ambiguity) in their ability to imitate may also view the downside to doing so sufficiently limited that they attempt it. This confidence may be misplaced. Even so, once the process of exploration begins, it may yet lead to success.

My results highlight three dimensions that are important in analyzing the sustainability of capabilities-based advantages under causal ambiguity: (1) the intrinsic riskiness of entry associated with the optimal structure, (2) intrinsic level of causal ambiguity of that structure, and (3) the accuracy of challenger beliefs with respect to the elements in this class. The first item is intimately related to the second. However, items 1 are 2 are not sufficient to guarantee a strong capability-based advantage—challengers' initial priors (item 3) also play a strong role. Sufficiently optimistic challengers always enter in the short term and may spend a long time experimenting, thereby reducing a strong capabilities-based advantage to, at best, a weak one.

If the implied performance differences between causal structures are small, then picking a random

[26] Lenox et al. (2006) is similarly careful; the interpretation of binary strings as "activity decisions" happens to arise naturally in their setting.

[27] At least to within an arbitrary margin of error.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

structure in the equivalence class is almost as good as imitating the incumbent. As described above, the bias under such circumstances should be toward sustained entry. However, there is also little incentive to literally imitate the incumbent, especially if changing causal relations involve switching costs (not considered here). Entry occurs and erodes the incumbent's profits, but the incremental return to entrants getting it exactly right is low. If imitation is risky (high intrinsic ambiguity coupled with high risk), the opposite dynamic is at work. That is, the risks of entry keep such activity low, but those firms that do enter are compelled to get it right.

One of the more interesting findings, especially given the growing interest in the relationship of complexity to performance, is that denser causal relationships do not necessarily imply greater causal ambiguity. The fact that more interrelationships between observable operating variables may actually reveal a lot to potential imitators has implications. Much has been written, for example, about the relative performance advantages enjoyed by Southwest Airlines and the difficulty of its larger competitors in their attempts to imitate it. This is true even though its activities are simpler than its competitors and fairly transparent (and, indeed, well documented). The organization is informal, there is no ticketing, routes are simple point-to-point, equipment is standardized, and teams are independent.[28] Each of these operational features imply either fewer influence relations between activities or greater difficulty in observing heterogeneity in outcomes (e.g., due to equipment standardization). These devices make strategic sense if they have the effect of simplifying operations in a way that increases the intrinsic ambiguity of Southwest's operations and, thereby, prevent imitation. Moreover, this argument does not rely upon any assumption of resource "stickiness" on the part of Southwest's competitors. Rather, the linkages adopted by Southwest may be sufficiently ambiguous, and the risk of experimentation sufficiently high, that imitation is foreclosed.

For similar reasons, small startups may be more difficult to imitate than larger, established firms—which may be one reason they tend to be good acquisition candidates (i.e., because this may be the only way to observe their hidden structure). Alternatively, outsourcing is a way to introduce independencies in observed operations. Done appropriately, this can actually increase the level causal ambiguity connected to a firm's strategy. So-called "flat" organizations (a feature of Southwest) also push in the direction of fewer interrelationships. Although it may be simple enough to create a flat organization, it may be quite difficult to do so successfully—even when observing the behavior of those who have.

Finally, let us speculate on some possible extensions of this theory. Throughout the analysis, it was assumed that the incumbent simply implemented the optimal structure. An obvious extension is to examine the incumbent as a strategic player. For example, it seems unlikely that a firm would adopt an easily imitated operating structure, even though highly efficient. What are the competitive trade-offs between operating performance and causal ambiguity? Answering this may improve our general understanding of when, e.g., the kinds of simplifying and decoupling devices seen at Southwest are likely to be implemented for strategic purposes.

It was also assumed that causal relationships between key, observable operating variables could be represented by directed acyclic graphs. When important operating variables are not observed, this assumption is no longer appropriate because correlations may be induced by hidden causes. In such cases, more general approaches must be used (e.g., "chain" graphs) to represent the relations observed by outsiders. Fortunately, the literature on probabilistic networks includes numerous approaches to this issue.

Another significant assumption was that managers knew the causal implications of their operating plans. However, the strategy literature also raises the possibility that these implications may not be known, resulting in causal ambiguity with respect to one's *own* structure. Also, the actions of senior managers aim not only to implement an appropriate set of interrelationships between operating entities but also to affect the local behaviors of those entities (i.e., to influence the parameters that determine the empirical distribution). It may be worthwhile to extend the analysis presented here to these cases.

Finally, much of the literature on probabilistic networks is concerned with the empirical exercise of estimating equivalence classes of causal structures from real-world history. This raises the interesting possibility of investigating the propositions presented here via empirical methods.

11. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

Acknowledgments

The author thanks R. Cassadesus-Masanel, D. Chickering, G. Cooper, J. Denrell, J. Gans, B. Gibbs, P. Ghemawat, H. Harmon, A. W. King, M. Lenox, D. Levinthal, G. MacDonald, J. Penalva-Zuasti, J. Rivkin, S. Rockhart, a remarkably patient and helpful anonymous referee, and participants at the 2006 Atlanta Competitive Advantage

[28] Saloner et al. (2001, pp. 67–71).

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

Conference and Academy of Management Conference, the strategy workshops at Carlos III University, Duke University, IESE, London Business School, University of Maryland, Rice University, University of Texas at Austin, and the Economic Theory workshop at New York University.

References

Barney, J. 1991. Firm resources and sustained competitive advantage. *J. Management* **17**(1) 99–120.

Besanko, D., D. Dranove, M. Shanley. 1996. *Economics of Strategy*. John Wiley & Sons, New York.

Besanko, D., U. Doraszelski, Y. Kryukov, M. Satterthwaite. 2007. Learning-by-doing, organizational forgetting, and industry dynamics. Harvard Institute of Economic Research Discussion Paper 2128. http://ssrn.com/abstract=962878.

Cohen, W. M., D. A. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Admin. Sci. Quart.* **35** 128–152.

Collis, D. J., C. A. Montgomery. 1998. *Corporate Strategy: A Resource-Based Approach*. McGraw-Hill, Boston.

Cowell, R. G., A. P. Dawid, S. L. Lauritzen, D. J. Spiegelhalter. 1999. *Probabilistic Networks and Expert Systems*. Springer, New York.

DeGroot, M. H. 1970. *Optimal Statistical Decisions*. McGraw-Hill, New York.

Dierickx, I., K. Cool. 1989. Asset stock accumulation and sustainability of competitive advantage. *Management Sci.* **35**(12) 1504–1511.

Edwards, D. 1995. *Introduction to Graphical Modelling*. Springer-Verlag, New York.

Einhorn, H. J., R. M. Hogarth. 1986. Decision making under ambiguity. *J. Bus.* **59** S225–S250.

Ghemawat, P. 1985. Building strategy on the experience curve. *Harvard Bus. Rev.* **63**(March–April) 143–149.

Ghemawat, P., D. Levinthal. 2000. Choice structures and business strategy. Working paper, Harvard Business School, Boston.

Golan, A., G. Judge, D. Miller. 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley & Sons, New York.

Grant, R. M. 2002. *Contemporary Strategy Analysis: Concepts, Techniques, Applications*. Blackwell, Malden, MA.

Jensen, F. V. 1996. *An Introduction to Bayesian Networks*. Springer, New York.

Jensen, F. V. 2001. *Bayesian Networks and Decision Graphs*. Springer, New York.

Kalai, E., E. Lehrer. 1993. Rational learning leads to Nash equilibrium. *Econometrica* **61**(5) 1019–1045.

Kauffman, S. A. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, Oxford, UK.

King, A. W., C. P. Zeithaml. 2001. Competencies and firm performance: Examining the causal ambiguity paradox. *Strategic Management J.* **22** 75–99.

Korb, K. B., A. E. Nicholson. 2004. *Bayesian Artificial Intelligence*. Chapman and Hall/CRC, Boca Raton, FL.

Lieberman, M. 1984. The learning curve and pricing in the chemical processing industries. *RAND J. Econom.* **15**(2) 213–228.

Lenox, M. J., S. F. Rockart, A. Y. Lewin. 2006. Interdependency, competition, and the distribution of firm and industry profits. *Management Sci.* **52**(5) 757–772.

Lenox, M. J., S. F. Rockart, A. Y. Lewin. 2007. Interdependency, competition, and industry dynamics. *Management Sci.* **53**(4) 599–615.

Levinthal, D. A. 1997. Adaptation on rugged landscapes. *Management Sci.* **43**(7) 934–950.

Levinthal, D. A., J. G. March. 1993. The myopia of learning. *Strategic Management J.* **14** 95–112.

Lippman, S. A., R. P. Rumelt. 1982. Uncertain imitability: An analysis of inter-firm differences in efficiency under competition. *Bell J. Econom.* **13**(3) 418–438.

March, J. G. 1991. Exploration and exploitation in organizational learning. *Organ. Sci.* **2**(1) 71–87.

Meek, C. 1995. Strong completeness and faithfulness in Bayesian networks. P. Besnard, S. Hanks, eds. *Proc. 11th Conf. on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, San Francisco, 411–418.

Moody, F. 1995. *I Sing the Body Electronic: A Year with Microsoft on the Multimedia Frontier*. Penguin Books, New York.

Mosakowski, E. 1997. Strategy making under causal ambiguity: Conceptual issues and empirical evidence. *Organ. Sci.* **8**(4) 414–442.

Nelson, R., S. Winter. 1982. *An Evolutionary Theory of Economic Change*. Belkap, Cambridge, MA.

Neopolitan, R. E. 2004. *Learning Bayesian Networks*. Pearson/Prentice Hall, Upper Saddle River, NJ.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. North Holland, Amsterdam.

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

Peteraf, M. A. 1993. The cornerstones of competitive advantage: A resource-based view. *Strategic Management J.* **14** 179–191.

Porter, M. E. 1980. *Competitive Strategy*. Free Press, New York.

Porter, M. E. 1996. What is strategy? *Harvard Bus. Rev.* **74**(6) 61–78.

Reed, R., R. J. DeFillippi. 1990. Causal ambiguity, barriers to imitation, and sustainable competitive advantage. *Acad. Management Rev.* **15**(1) 88–102.

Rivkin, J. W. 2000. Imitation of complex strategies. *Management Sci.* **46**(6) 824–844.

Robinson, R. W. 1977. Counting unlabeled acyclic digraphs. C. H. C. Little, ed. *Lecture Notes in Mathematics, 622: Combinatorial Mathematics V*. Springer-Verlag, New York.

Ryall, M. D. 2003. Subjective rationality, self-confirming equilibrium, and corporate strategy. *Management Sci.* **49**(7) 936–949.

Saloner, G., A. Shepard, J. Podolny. 2001. *Strategic Management*. John Wiley & Sons, New York.

Spirtes, P., C. Glymour, R. Scheines. 2000. *Causation, Prediction and Search*. MIT Press, Cambridge, MA.

Tirole, J. 1988. *The Theory of Industrial Organization*. MIT Press, Cambridge, MA.

Verma, T., J. Pearl. 1990. Equivalence and synthesis of causal models. M. Henrion, L. N. Kanal, J. F. Lemmer, eds. *Proc. 6th Conf. on Uncertainty in Artificial Intelligence*. Elsevier, New York, 255–270.