The Fog of Change: Opacity and Asperity in Organizations

Michael T. Hannan
Stanford University
László Pólos
Erasmus University, Rotterdam and
Loránd Eötvös University, Budapest
Glenn R. Carroll
Stanford University

Initial architectural change in organizations often induces other subsequent changes, generating lengthy cascades of changes in subordinate units. This article extends a formal model of cascading organizational change by examining the implications for organizational change of the limited foresight of those who initiate such change about unit interconnections (structural opacity) and the normative restrictiveness imposed on architectural features by organizational culture (cultural asperity). Opacity leads actors to underestimate the lengths of periods of reorganization and the associated costs of change, thereby prompting them unwittingly to undertake changes with adverse consequences. Increased opacity and asperity lengthen the total time that the organization spends reorganizing and the associated opportunity costs; and the expected effect of an architectural change on mortality hazards increases with the intricacy of the organizational design, structural opacity, and the asperity of organizational culture. We illustrate the theory with an interpretation of the 1995 collapse of Baring Brothers Bank.●

Understanding organizational change, particularly in internal structures, remains a core issue in organizational theory, with all contemporary theoretical perspectives offering insights and arguments. Most recently, Hannan, Pólos, and Carroll (2003a) have developed a theory of change in organizational architecture, depicted as a code system, that represents change as a cascading process: an initial architectural change induces other changes in the organization, generating a cascade of changes. The main theoretical argument ties two organizational properties to the total time that the organization spends reorganizing and to the associated opportunity costs. Specifically, the expected deleterious effect of a change in an organization's architecture on its mortality hazard increases with intricacy (interconnectedness among units of the organizational design) and viscosity (sluggishness of response). But the expected effects of an architectural change might be complicated by other factors.

A key question about architectural change concerns the impetus for an initiating change event: Why would any organization undertake a change that was likely to take a long time to complete and thereby cause it to miss many opportunities and risk failure? A common answer in organizational theory is that such outcomes are unanticipated because bounded rationality leads decision makers to miscalculate the costs and risks relative to the expected gains (March and Simon, 1958; Cyert and March, 1963). This is a sound explanation, but there are several specific ways that bounded rationality might manifest itself in initiating architectural change that looks good and might produce a positive outcome yet can also generate potentially dire results. Two specific factors that might affect the consequences of architectural change are (1) structural limits on foresight of those initiating a change and (2) cultural opposition to the architectural change.

Limited foresight produces a systematic tendency to underestimate the length of reorganization periods and thus to underestimate the costs of change. Given such systematic

0001-8392/03/48-03-0399/$3.00.

●

This research was supported by fellowships from the Netherlands Institute for Advanced Study, the Stanford Graduate School of Business Trust, ERIM at Erasmus University, and the Centre for Formal Studies in the Social Sciences at Eötvös University. We benefited from the comments of Jim Baron, Dave Barron, Jerker Denrell, Gábor Péli, the participants in the workshop of the Nagymaros Group on Organizational Ecology and in the Stanford Strategy Conference, and the *ASQ* referees.

underestimation, organizational leaders can easily choose to enter into changes that cost far more than the expected benefits of successfully completing the change. For example, analysts generally agree that in 1999 the management of Xerox Corporation underestimated the difficulty of a transformation it undertook in an attempt to improve its cost structure and provide better service. At a time when the company was doing very well by most observable measures—increasing profits, high stock price, and no high-end competitors with comparable products—Xerox Corporation simultaneously reorganized the architectures of its sales and billing functions. The billing reorganization consolidated 36 administrative centers into three. The sales reorganization shifted its staff of 15,000 persons from units based on geography to those based on industry, transforming positions from local generalists to national product specialists. The results proved disastrous: billing errors proliferated, and sales staff spent much of their time resolving problems rather than learning their new roles and making contacts with buyers that they had never met. Staff turnover rose, sales dropped, and customers moved to competitors. At the same time, Xerox faced stiff new competition on its previously unrivaled high-end copiers from both Heidelberger Druckmaschinen AG and Canon Inc. Within 18 months the losses had become so substantial that a recently installed chief executive officer (CEO) was ousted, and the business press speculated that the company would not survive. As the *Wall Street Journal* noted, "retraining to sell and service such intricate machines proved more difficult than the company anticipated" (Pereira and Klein, 2000: A3).

Cultural opposition might affect cascades of architectural change because it signifies a shift in the meaning of a proposed architectural change, turning what was likely viewed initially as a dispassionate cost-benefit calculation into a normative matter. In Selznick's (1948) words, these architectural features have become infused with moral value. The strength of cultural opposition can be difficult to anticipate, and the turmoil associated with it lengthens the reorganization period. For example, at Apple Computer in 1997, CEO Gilbert Amelio introduced a centralized architecture to a culture that "always championed the individual and stressed freedom to act unilaterally" (Amelio, 1998: 228). Amid much strife, Apple's transformation continued down a stormy path until Amelio was ousted and replaced by the founder, Steve Jobs.

In examining limited foresight and cultural opposition, we extend Hannan, Pólos, and Carroll's (2003a) model of cascading organizational change to include these ideas. Specifically, we introduce and develop two concepts: opacity, defined as limited foresight about interconnections among an organization's units, and asperity, defined as normative restrictions on certain architectural features. Opacity leads actors to underestimate the length of reorganization and the associated costs of change, thereby prompting them to undertake changes with adverse consequences. Both opacity and asperity increase the time spent reorganizing and raise the opportunity costs of change and the expected effect of architectural change on mortality. To make connections with other theoretical developments clear, we build formal representations of the theoretical arguments, using a formal structure that eases integration of the various strands of theory on structural inertia and change: the nonmonotonic logic developed by Pólos and Hannan (2001, 2002, 2004). We illustrate the concepts and arguments by reinterpreting the circumstances surrounding the spectacular collapse of Barings Bank in 1995, which involved both limited foresight and cultural opposition, in our interpretation.

ORGANIZATIONAL ARCHITECTURE AND CULTURE

Organizational analysts commonly distinguish between formal architectural and informal cultural features. Architecture refers to the formal (official) specifications of an organization and its governance. Architectural choices are reflected in the formal structures for assigning work, that is, constructing the units that undertake the subtransactions. The choices also specify the means of coordinating members and units, monitoring them, and allocating resources and rewards. Culture governs how work actually gets completed, how members interact, how decisions are actually made, which units defer to others, and so forth. The notion of culture includes both tacit knowledge of the details of the work process, including locally generated knowledge and craft/professional knowledge generated outside the organization, and norms encoding the informal understandings and practices for interaction, authority, and so forth. When viewed abstractly, specific architectural and cultural elements can be regarded as the values of functions that specify organizational features. As Simon (1954) explained in analyzing the employment relation, the values of such feature values are not defined with absolute clarity; rather, they allow a certain amount of tolerance. Simon postulated that the employment relation is such that the employee allows the employer to assign tasks from some restricted range of options. We posit that architectures and cultures discriminate between the allowed and disallowed feature values, thereby imposing constraints on feature values and limiting the values that they can legitimately take.

An appropriate language for expressing architectures and cultures should reflect these considerations. Moreover, it should allow precise judgments about the consistency of architectures and cultures. We use a semantic formulation, representing architectures and cultures as collections of sentences pertaining to ontology (e.g., definitions of the units in an architecture) and rules (e.g., statements of which units have authority over which other units). Such sentences form a code, which can be understood as both (1) a specification of a blueprint, as in the genetic code, and (2) a rule of conduct, as in the penal code (Pólos, Hannan, and Carroll, 2002). Our use of the term code reflects both meanings.

Some codes matter greatly in the sense that violations are punished very severely while others are handled with a lighter touch, as with the distinction between felonies and misdemeanors. Modeling the severity of codes poses numerous challenges, and we do not attempt it here. Instead, we restrict the theory to apply to serious codes. Henceforth, when we refer to architectural and cultural codes, we mean only the serious ones, those whose observable violation brings strong punishments.

To represent organizational architectural and cultural codes formally, we start by introducing some primitive terms. First, not all kinds of organizations can be compared straightforwardly. Comparisons of organizations make the most sense when restricted to populations of organizations (Carroll and Hannan, 2000). We link the arguments to populations by using the two-place predicate O*(o,p)* that reads as "*o* is an organization in organizational population *p*." We intend that the theory be understood as applying to all populations of organizations, and we do not know of any exceptions. If we did know of such exceptional cases, we would express the entire theory as holding as a rule with exceptions, and we would employ nonmonotonic quantification over populations. In terms of the formalisms of the theory-building strategy, this would work as follows. Suppose that $\varphi(p)$ is a postulate of the theory. Then, the theoretical claim formulated in a way that allows for accidental exceptions would be stated as $\mathfrak{N}p[\varphi(p)]$. (We explain the quantifier $\mathfrak{N}$ below.) We do not introduce this level of complication in our formal rendering of the theory.

Our argument will specify some basic processes at the level of the organization unit. The predicate $\mathrm{U}(u,o)$ says that "*u* is a unit of the entity *o*, where $\mathrm{O}(o,p)$." We add the background assumption that units belong to only one organization. In formal terms,

$$\forall o,o',u[\mathrm{U}(u,o) \wedge \mathrm{U}(u,o') \rightarrow (o = o')].$$

Tables 1 and 2 summarize our notation.

An organizational architecture can be regarded as a set of values of features, such as the form of authority, pattern of control relations, accounting principles, compensation policies, and so forth. Some relevant features that pertain to its global architecture are common to units in an organization; others vary by type of unit. We describe the architecture of a unit by identifying the relevant features and determining the set of feasible alternative values of each feature. For instance, the form of authority might be a relevant feature; and the relevant alternatives might be "bureaucratic," "professional," and "charismatic." In other words, we consider features to be functions that map from organizations and time points to the range of possible values. We denote the kth feature of unit *u* at time *t* as $f_k(u,t)$, and we denote the range of possible values by (the set) $\mathsf{A}_k(u,t)$. The Cartesian product of the sets of possible values taken over all of the relevant features: $\mathsf{A}_{ut} \equiv \mathsf{A}_{1ut} \times \mathsf{A}_{2ut} \times \cdots \times \mathsf{A}_{Kut}$ gives the space of potential architectures for the unit. Finally, we let $\mathbf{a}_{ut}$ denote the unit's actual architecture, the set of choices of values for each of the relevant features. When we want to instantiate formally that $\mathbf{a}_{ut}$ is a unit's actual architecture, we use the predicate $\mathrm{ARCH}(u,t,\mathbf{a}_{ut})$, which reads as "$\mathbf{a}_{ut}$ is the K-tuple of feature values that specifies unit *u*'s architecture at time *t*."

Table 1

Notation for Logical Constants, Predicates, Functions, and Relations

Logical constants	
$\vee$	Disjunction
$\wedge$	Conjunction
$\neg$	Negation
$\exists$	Classical existential quantifier
$\forall$	Classical universal quantifier
$\rightarrow$	Classical material implication
$\mathfrak{N}$	Nonmonotonic "normally" quantifier
$\mathfrak{A}$	Nonmonotonic "auxiliary" quantifier
$\mathfrak{P}$	Nonmonotonic "presumably" quantifier
ARCH(u,t,α)	Unit u's architecture at t is given by the codes α
CUL(u,t,γ)	Unit u's culture at t is given by the codes γ
O(o,p)	o is an organization in population p
OP(u,u')	Unit u' is opaque to unit u
RE(u,t)	Unit u is in reorganization at time t
U(u,o)	Unit u is a unit in organization o
Functions	
A_o	Asperity of organization o
$c(u)$	Centrality of unit u
I_o	Intricacy of organization o
I_o^*	Foreseen intricacy of organization o
P_o	Opacity of organization o
Relations	
$\succ$	Dominance relation over units

Table 2

Notation for Random Variables, Probabilities, and Key Parameters

Random variables	
$d(u,t)$	Duration of an induced violation in unit u
$D(o,t \mid \delta(u,t) = 1)$	Sum of the durations of the induced violations in a cascade initiated by unit u at time t
$D(o,t)$	Sum of the durations of the induced violations in a cascade initiated by a random unit at time t
$\delta(u,t) = 1$	Unit u initiates an architectural code change just before t (= 0 otherwise)
$\Delta(o,t) = 1$	Organization o experiences an architectural code change just before t (= 0 otherwise)
$m(u,w,w')$	Number of opportunities missed by unit u during $[w,w')$
$M(u,w,w')$	Number of opportunities missed by organization o during $[w,w')$
$\mu(o,t)$	Organization o's hazard of mortality at t
$N(o,t \mid \delta(u,t) = 1)$	Number of units with induced violations in a cascade initiated by unit u at time t
$N(o,t)$	Number of units with induced violations in a cascade initiated by a random unit at time t
$R(o,t)$	Organization o's stock of resources at time t
$S(o,t)$	Temporal span of a cascade in organization o starting at t
$v(u,u',t) = 1$	Unit u induces an architectural code violation in unit u' at t (= 0 otherwise)
λ_o	Hazard of initiating architectural change for units in o
π_o	Probability of induced arch. code violation for units in o
Population-specific parameters	
η_o	Probability that a unit misses an opportunity while not in reorganization
$\tilde{\eta}_o$	Probability that a unit misses an opportunity while in reorganization
ξ_o	Arrival rate of opportunities for a unit in organization o
Ξ_o	Arrival rate of opportunities for organization o

Architectural codes restrict the set of allowable architectures. A sharp architectural code rules out many of the possibilities in A_{ut}; a loose architecture places few constraints on the architectural choices of the unit. It is important to distinguish architectural codes that lie under the control of the unit from those imposed externally. Let $\alpha_{ut} \subseteq \mathrm{A}_{ut}$ denote the set that contains the officially approved architectures for unit *u* at time *t*, and let α^{i}_{ut} and α^{e}_{ut} be the internally and externally controlled subcodes, respectively. (For simplicity, we assume that the two subcodes are disjoint: $\alpha_{ut} = \alpha^{i}_{ut} \cup \alpha^{e}_{ut}$.) The imposed codes reflect a superordination relation among units; they could arise from specified lines of authority, from the flow of work, or from any similar relation that allows one part of the organization to impose constraints on another part. An important class of interunit relations concerns subordination in choice of architecture. Let $u \succ u'$ specify that unit "*u* and *u′* are units in the same organization and *u* is superordinate to unit *u′* in the sense that choices of architectural feature values by *u* create architectural code restrictions (binding constraints) for unit *u′*."

Culture

Cultural codes are typically implicit and local and govern the informal organization. They specify how things actually get done, how people interact, the bases of status, and so forth. Architectural and cultural codes generally differ with respect to enforcement. An agent in authority typically uses organizational authority to compel compliance with the rule. This means using organizational sanctions such as official reprimands, threats of demotion or budget cuts, and so forth. The enforcement mechanism for cultural codes works by more subtle means. Both insiders and outsiders hold expectations of the organization based on cultural codes, and they devalue organizations that violate applicable cultural codes. Cultural-code violations are often prevented simply by the threat of devaluation. Furthermore, persistent perceived code violations, in turn, produce a sequence of drops in valuations (for details on these processes, see Pólos, Hannan, and Carroll, 2002). Of course, the literature on organizations contains many stories and ideas about how organizations manage to continue to operate while experiencing code violations.

Externally enforced cultural codes reflect some kind of broad external constraint, such as codes and norms of professional conduct, regulations, and laws. Internally enforced codes constitute the local culture—traditions and stable expectations about various aspects of life within an organization. In this article, we restrict our attention to those aspects of the culture that bear directly on architectural choices. We assume that local culture generally sets range restrictions on architectures, producing strong resistance should violation be attempted. For example, when officials at Ben & Jerry's Ice Cream proposed to relax the rule that set a ceiling on the ratio of the highest to the lowest level of compensation in the firm, a strong cultural reaction opposed this architectural change (Lager, 1994). Similarly, the faculty cultures of most top research-oriented business schools are likely to strenuously resist attempts by a dean (say, a former chief executive officer) to impose command-and-control decision making for faculty personnel decisions.

We define the architecturally relevant local cultural code, γ_{ut}, as the subset of A_{ut} that is not forbidden by the cultural code of the unit. When we want to instantiate formally that γ_{ut} is (part of) a unit's cultural code, we use the predicate $\mathrm{CUL}(u,t,\gamma_{ut})$, which says that "$\gamma_{ut} \subseteq \mathrm{A}_{ut}$ is the set of architectures that satisfy unit u's cultural code at time t."

A restrictive culture excludes many architectural possibilities. We call the level of such cultural restrictiveness *asperity*, relying on the dictionary definition of asperity as "severity or rigor," which determines the normative restrictiveness on certain architectural features. In formal terms, we represent asperity as follows.

Definition 1 (D.1). The asperity of a unit's architecturally relevant culture equals the fraction of possible architectures that the culture rules out.

$$\text{Assume } \mathrm{ARCH}(u,t,\alpha_{ut}) \wedge \mathrm{CUL}(u,t,\gamma_{ut}).$$

$$a(u,t) = 1 - \frac{|\gamma_{ut}|}{|\alpha_{ut}|},\ |\alpha_{ut}| > 0,$$

where $|\cdot|$ denotes the cardinality of a set, the number of unique elements it contains. Note that this measure of asperity equals one if the culture does not admit any of the officially warranted architectures; it equals zero if the culture does not rule out any of the approved architectures.

THEORY STAGE 1: INTRICACY AND RANDOM CASCADES

We build on the model of random cascades developed by Hannan, Pólos, and Carroll (2003a), which we call Theory Stage 1 and briefly review here. This model concerns changes in some architectural features and/or codes by some particular organizational unit, which might sit anywhere in the organization's formal hierarchy. The reasons for the initial change are not pertinent to the theory; they could encompass a wide variety of possibilities, including changes in external opportunities and constraints, executive tinkering, and internal strife. The specific change undertaken might be sensible in that it would likely improve organizational alignment and functioning, but we do not assume this—changes can also degrade performance. Our focus on architectural changes as triggers of cascades reflects the view that architecture is generally more malleable to management and to individual decision makers: changing the architecture often requires only a directive to do so from someone with authority.

Let the random variable $\delta(u,t)$ equal one if unit u changes its architecture by replacing an architectural code or a feature value just after t; and it equals zero otherwise. When the reference is to an organization, rather than a unit, undergoing architectural change, the random variable $\Delta(o,t)$ is used; it equals one if any unit in organization o experiences architectural change at (just after) time t, and it equals zero otherwise.

$$\Delta(o,t) = 1 \leftrightarrow \exists u[\mathrm{O}(o,p) \wedge \mathrm{U}(u,o) \wedge \delta(u,t) = 1)].$$

Inconsistencies between new and existing codes normally become salient and consequential when actions that would have satisfied the old architectural code do not satisfy the new one. We represent this formally by defining violations of a unit's architectural code that are induced by another unit. The random variable $v(u,u',t)$ equals one if unit u' induces an architectural code violation for unit u at time t and equals zero otherwise. Once a unit experiences an induced violation, the violation persists until it adjusts its architecture to conform to the newly imposed constraints, though we do not assume that all violations get fixed; the duration might be very long. We denote the time that passes between the induction and the elimination of an architectural code violation, the *duration* of the induced violation, with the random variable $d(u,t)$. A key step in building the model is defining a cascade of induced changes initiated by an original (uninduced) change in architecture by a unit in the organization.

Definition 2 (D.2). A particular cascade of resolutions of induced architectural-code violations in organization o that begins at time t with a change initiated by unit u is constructed as follows:

Step 0. The unit u, not in violation of any of its applicable architectural code, initiates the cascade at time t by changing its architectural code, and that change induces architectural-code violations in one or more other units;

Step 1. A unit with an induced violation in step 0 changes its architecture such that conformity eliminates the induced violation, but this architectural change induces a violation in one or more other units;

⋮

Step L. The only unit with an unresolved induced violation (generated by the previous steps in the cascade) eliminates the violation at time t_L, and this architectural change does not induce a violation in any unit.

We denote the set of induced changes such a cascade comprises as $\mathrm{K}(u,t)$, where the variables identify the unit that initiated the cascade (u) and the time of initiation (t).

We pay special attention to the temporal dimensions of a random cascade in formulating substantive ideas such as those pertaining to cultural resistance. We see two different ways to think about the longevity of such cascades. One is to consider the *temporal span*, $S(o,t)$, the time elapsed from the initiating event to the elimination event that terminates the cascade. The other is to examine the total amount of time spent by units in reorganization, changing codes and feature values so as to eliminate induced violations. The *total time in reorganization* during a cascade, $D(o,t)$, is calculated by summing up the times taken in all the individual stages of the cascade, even if they occur simultaneously. Both concepts have substantive promise, because a protracted period of reorganization presumably complicates organizational action and diverts the attention of at least some members over the whole period—the first idea—and because the disruption caused by reorganization ought to be proportional to the time spent by units (and their members) in working out the consequences of changes—the second idea. Because we model changes at the unit level, it will simplify the analysis to concentrate on total reorganization time.

The Probability Model for Random Cascades (Π)

The first stage of the theory represents expected outcomes in cascades initiated by a unit chosen at random. The treatment of cascades with random initiators is crucial for representing the kinds of comparisons made in empirical research on the consequences of structural change, which rarely has access to the full history of all of the cascades of change. As we explain in the Appendix, we refer to the simplifying assumptions that we expect to see relaxed in further developments of the theory as auxiliary assumptions, and we refer to the real causal claims of the theory as postulates. In this analysis, the auxiliary assumptions concern the probabilistic structure of the change process. We quantify auxiliary assumptions using the $\mathfrak{A}$ quantifier discussed in the Appendix. The probability model used for this purpose, which we denote by Π, makes the following auxiliary assumptions:

1. The probability that a change in a unit induces a violation in a subordinate unit, $\Pr\{v(u,u',t)\}$, does not vary between pairs of units or over time; it is an organization-specific constant, π_o.
2. The expected time to elimination of an induced violation of architectural code does not vary over units or over time; it is an organization-specific constant: τ_o, which gives the organization's viscosity or sluggishness in response.
3. The hazard of initiating an architectural change does not vary over units or over time; it is an organization-specific constant: $\lambda_o > 0$.
4. The arrival rate of opportunities for the organizations in a population varies over time but does not vary over units at a point in time: $\Xi > 0$; the arrival rate for a unit in an organization is Ξ / U_o, where U_o equals the number of units in organization o.
5. The probability that a unit misses an opportunity while in reorganization and the probability of missing an opportunity in the absence of a reorganization does not vary over time or among units in the organizations in a population; they are population parameters, $\eta + \tilde{\eta}$ and η, respectively ($\eta, \tilde{\eta} > 0$).

Two units are connected in an architectural sense if the feature values of one govern and constrain the architectural codes of the other. In particular, one unit constrains another architecturally if the feature values and choices allowed by the codes of the former are imposed as an external constraint (as codes) on the latter. Let U_o give the number of units in an organization. Consider a $U_o \times U_o$ adjacency matrix **R** for which r_{ij} = 1 if unit *i* constrains unit *j* architecturally and equals 0 otherwise, that is, if $u_i \succ u_j$. Furthermore, let *c(u)* be a function that gives the architectural centrality of unit *u*. The vector of centrality scores of the units of organization *o* at time *t* is given by a slight variation on Bonacich's (1987) measure of centrality:

$$\mathbf{c} = \sum_{k=1}^{\infty} (\pi_o)^k \, \mathbf{R}^k \, \mathbf{I},$$

where **I** is a (N×1) vector of ones.

Let $N(u,t \mid \delta(u,t) = 1)$ be a random variable that records the number of induced violations in the cascade $\langle u', t' \rangle \in \mathbf{K}(u,t)$. Under the probability model Π, the architectural centrality of the unit that initiates a cascade gives the expected number of induced violations in the cascade. This is the key to the model. In expressing this result formally, we use the following notation to denote the operation of mathematical expectation: $\mathrm{E}\{\cdot\}$. We also introduce a shorthand expression for the sum of the durations of all reorganizations triggered by the cascade $\mathbf{K}(u,t)$:

$$D(u,t \mid \delta(u,t) = 1) \equiv \Sigma_{\langle u', t' \rangle \in \mathbf{K}(u,t)} d(u', t').$$

Lemma 1 (L.1). The expected number of induced violations resulting from an architectural change by a unit equals its centrality in the architecture, $c(u)$ (L.1 in Hannan, Pólos, and Carroll, 2003a):

$$\mathfrak{P} o,u,t[\mathrm{O}(o,p) \wedge \mathrm{U}(u,o) \rightarrow \mathrm{E}\{N(o,t \mid \delta(u,t)) = 1)\} = c(u)].$$

This formula says that a minimal rule chain in this stage of the theory supports the claim that the expected number of induced violations in a cascade with random origin equals the centrality of the initiating unit (the proof is given in Hannan, Pólos, and Carroll, 2003a). It should be noted that the generalizations pertain to properties of *probability distributions*. An exception to these generalizations would be a population for which the probability distribution deviates from the normal pattern, not stochastic variation within a population governed by a particular distribution.

The next step in representing random cascades allows the origin (the unit initiating the cascade) to be chosen at random so that changes can be compared across organizations in a population. The mean centrality score in an organization provides a useful way to express intuitions about likely lengths of such random cascades. For a unit to have a high centrality score, it must dominate units that are themselves high in centrality. Therefore, cascades are more likely to hit units with high centralization in an organization with a high mean centrality. Mean centrality provides a way to characterize the *intricacy* of the organization's design; *The Oxford English Dic-*

tionary defines intricate as "perplexingly entangled or involved; interwinding in a complicated manner." Let I_o denote a non-negative function that records the intricacy of the design of organization *o*.

Definition 3 (D.3). The intricacy of an organization's design equals the mean of the centralities of its subunits:

$$I_o \equiv (1/U_o) \sum_{u=1}^{U_o} c(u).$$

Analyses of cascades with a random initiating unit involve calculating expectations of functions of random cascades. These functions involve summations over all of the induced violations in a cascade. In the standard case, in which the size of the set of elements in the summation is deterministic, the straightforward calculation uses the rule that the expectation of a sum of functions of random variables is the sum of the expectations. In the case of cascades, the number of elements in the summation is itself a random variable, $N(o,t \mid \delta(u,t) = 1)$ if the cascade initiates in unit *u* as we noted above. Recall also, according to L.1, that $\mathrm{E}\{N(o,t \mid \delta(u,t) = 1)\} = c(u)$.

Now consider the case of an initiating unit chosen at random, as specified in the auxiliary assumptions. Let the number of induced violations in a cascade with random origin be denoted $N(o,t)$, the time spent reorganizing in a random case as $D(o,t)$, and the temporal span of a cascade, the time elapsed from origin to conclusion, by $S(o,t)$; the absence of the condition that identifies the unit initiating the cascade distinguishes these random variables from those introduced earlier.

Lemma 2 (L.2). The expected number of induced violations within a cascade with a random origin equals the organization's intricacy (L.2 in Hannan, Pólos, and Carroll, 2003a):

$$\mathfrak{P}\, o,u,t[\mathrm{O}(o,p) \wedge \mathrm{U}(u,o) \wedge (\Delta(o,t) = 1) \rightarrow \mathrm{E}\{N(o,t)\} = I_o].$$

The rest of Stage 1 of the theory derives implications of cascades for an organization's life chances using arguments about the effect of reorganization on the probability of missing opportunities and the effect of missing opportunities on the growth or decline in resources. It generates a predicted rise in mortality hazards for organizations with high levels of intricacy because of the opportunities they miss while reorganizing (an empirical pattern different from that assumed by Barnett and Carroll, 1995). We do not restate this part of the theory here. Some of it is overridden by the more specific arguments that we present below. The enduring part will be restated as the flow of the argument demands.

LIMITED FORESIGHT: STRUCTURAL OPACITY

The theory of Stage 1 provides a tool for analyzing organizational change; it explains how certain architectural changes in certain contexts might increase an organization's chances of failure. But recognizing such possibilities raises the question of why any organization would initiate such a hazardous change. We assume that managers are intendedly rational and well-meaning, that they are not intentionally destructive when they initiate such changes but, instead, believe they are making organizational adjustments that will improve performance and life chances. So the initial change must involve a miscalculation that results in an unexpected outcome (Kahneman and Lovallo, 1993; Sterman, Repenning, and Kofman, 1997).

Although many possible forms of miscalculation could produce this result, we focus on one that is only subtly connected to the outcome and that can be exploited in the model. We concentrate on the opacity of some of the interconnections among units in the organization and the limited foresight that results. In other words, we assume that reduced foresight about the exact structure of connections among units impairs the ability to forecast accurately the costs and benefits of a change. Such miscalculation can give the organization a rosier-than-justified expectation and thus prompt it to undertake changes with deleterious results.

We distinguish what is foreseen when reorganization begins from what actually happens. We mark predicates and functions that refer to foresight with an asterisk; those that refer to the facts, as they emerge, are not so marked. Using our notation for foresight, we define $N^*(o,t \mid \delta(u,t) = 1)$ as the random variable that records the number of induced violations of new code resulting from a cascade initiated by unit *u* that can be foreseen by the relevant actors in that unit at time *t*; and $D^*(u,t \mid \delta(u,t) = 1)$ as the random variable that records the foreseen length of the cascade from the vantage of unit *u* in organization *o* at time *t*.

We assume that actual periods of reorganization lengthen to the extent that the number of induced violations exceeds the limits on foresight. When violations cannot be foreseen, agents cannot plan comprehensively for reorganization and cannot undertake as many adjustments in parallel. The fact that the unforeseen architectural code violations show up in mid-change slows the process of reorganization, thereby extending actual periods of reorganization.

In many cases, an organization's structure imposes limits on what can be known, with the result that information about some parts of an organization is unavailable in other parts, what Williamson (1975) called information impactedness. For instance, Stinchcombe (1990: 75) observed that local unit information usually does not "have to flow anywhere in a hierarchy, except in very aggregated form as a budget estimate." Sometimes the languages used in different parts of the organization differ such that those outside the unit cannot interpret a full-disclosure description of the activities in an organizational unit. Other times, lack of transparency arises due to strategic withholding of information.

Stinchcombe (1990: 81) noted that pools of local knowledge might facilitate cultural developments that heighten opacity:

> When a large share of the information used in a given activity is such local knowledge, a subculture grows that is more or less isolat-

ed from the rest of the organization. The subculture is organized in a large measure around the information system that is of little use or interest to anyone else and so is adapted to particular concrete features, . . . uses an arcane language or system of notation and resists invasion by standards from larger and more uniform information systems.

Although opacity might have many potential meanings in an organizational context, we apply the idea here in a very specific sense. Specifically, one unit is structurally opaque to another, as we define the term, if the connections that flow from the former cannot be seen by the latter. To represent opacity in terms of the formal model used to analyze random cascades, we treat the matrix **R**, which records the superordination relations among the units of the organization, as factual and compare it with the potentially clouded vision from the vantage points of the various units. We use the predicate OP(u,u'), which says as "unit u's vision of unit u' is opaque in the sense that u cannot see u'"s superordination relations." Then, we define the matrix $\mathbf{R}^*_{\boldsymbol{u}}$, the matrix of architectural relations in the organization from u's (potentially opaque) vantage point.

Definition 4 (D.4). $\mathbf{R}^*_{\boldsymbol{u}}$ is the $U_o \times U_o$ matrix whose (i,j)th element is given by

$$r^*_u(i,j) = \begin{cases} 1 & \text{if } r(i,j) = 1 \wedge \neg \mathrm{OP}(u,u_i); \\ 0 & \text{otherwise.} \end{cases}$$

Note that this definition imposes the constraint that the u'th row in $\mathbf{R}^*_{\boldsymbol{u}}$ is set to zero if u' is opaque to u. This means that the dominance relations of u' over other units are obscured.

We can see the consequences of opacity by considering the centrality scores based on $\mathbf{R}^*_{\boldsymbol{u}}$.

Definition 5 (D.5). The partially obscured centrality of unit u at time t is given by

$$c^*(u) = \mathbf{b}\sum_{k=1}^{\infty} \pi_o^k (\mathbf{R}^*_{\boldsymbol{u}})^k \, \mathbf{I},$$

where **b** is an $(1 \times U_o)$ vector with $b_i = 0$ for $i \neq u$ and $b_i = 1$ for $i = u$. (Note that premultiplying by **b** simply picks out the uth element in the vector of partial centrality scores from the vantage point of u.) We refer to the difference between the actual and partially obscured centralities of a unit as the opacity of the structure from its perspective.

Definition 6 (D.6). The opacity of the structure for a unit is the difference between the unit's actual and foreseen centrality scores:

$$p(u) = c(u) - c^*(u).$$

With this definition in hand, we obtain a lemma paralleling L.1, which holds that the expected number of induced violations in a cascade with known origin equals the centrality of the initiating unit.

Lemma 3 (L.3). The expected number of induced violations in a cascade not foreseen from the vantage of the initiating unit equals the opacity of the initiator: $p(u)$.

$$\mathfrak{P} o,u,t[\mathrm{O}(o,p) \wedge \mathrm{U}(u,o) \rightarrow \\ \mathrm{E}\{N(o,t \mid \delta(u,t) = 1) - N^{*}(o,t \mid \delta(u,t) = 1)\} = p(u)].$$

Proof. In the nonmonotonic logic we use, establishing a proof involves constructing the most-specific regularity chains that connect the antecedent and the consequent. The regularity chains are constructed from the available definitions, postulates, strict rules, and auxiliary assumptions. If the most-specific such regularity chain supports the claim, then the theorem is proven. If among the most-specific regularity chains, some support the claim and others support the counter claim, then no conclusion is warranted—the claim is not a theorem. Therefore, we construct the most-specific regularity chains in sketching each proof.

The expected foreseen number of induced violations, $\mathrm{E}\{N^{*}(o,t \mid \delta(u,t) = 1)\}$, can be expressed as the sum of the expected foreseen number of violations at each path length. That is, $\mathrm{E}\{N^{*}(o,t \mid \delta(u,t) = 1)\} = \Sigma_{k=1}^{\infty} \mathrm{E}\{N^{*k}(u,t)\}$, where $\mathrm{E}\{N^{*k}(u,t)\}$ is the expected foreseen number of induced violations at step k of the cascade (that is, at path length k). The joint probability of k inductions along a path, under Π is given by π_{o}^{k}. Because inductions must follow the subordination relation, $\mathrm{E}\{N^{*k}(u,t)\} = \Sigma_{u \neq u'} \pi_{o}^{k}\, z^{*k}_{u,u'}$, where $z^{*k}_{u,u'}$ equals the number of distinct k-step foreseen paths connecting u and u'. Inspecting the terms in the powers of $\mathbf{R}^{*}$ reveals that $z^{*k}_{u,u'}$ is the (u,u') entry in $(\mathbf{R}^{*})^{k}$. Therefore, the expected foreseen number of induced violations equals $c^{*}(u)$. By L.1, the expected actual number equals $c(u)$. Use of D.6 completes the regularity chain.

Next we define a function, P_{o}, that defines opacity at the organizational level as the average of the opacities of its units.

Definition 7 (D.7). An organization's opacity equals the difference between the actual intricacy and the partially obscured intricacy:

$$P_{o} = (1/U_{o}) \sum_{u=1}^{U_{o}} p(u) = I_{o} - I_{o}^{*},$$

where $I_{o}^{*} = (1/U_{o}) \Sigma_{u=1}^{U_{o}} c^{*}(u)$.

Note that $P_{o} = 0$ if and only if the centralities are not obscured for any unit. We call this the case of full transparency. $I_{o} = 0$ logically implies that $P_{o} = 0$, because $I_{o}^{*} = 0$ in this case.

Lemma 4 (L.4). The expected number of unforeseen induced violations in a cascade initiated by a random unit equals the organization's opacity:

$$\mathfrak{P}o,t[\mathrm{O}(o,p) \wedge (\Delta(o,t) = 1) \rightarrow \mathrm{E}\{N(o,t) - N^*(o,t)\} = P_o].$$

Proof. By definition $\mathrm{E}\{N(o,t) - N^*(o,t)\} = \mathrm{E}\{N(o,t)\} - \mathrm{E}\{N^*(o,t)\}$. According to L.2, $\mathrm{E}\{N^*(o,t)\} = I_o$. By the law of total probability,

$$\mathrm{E}\{N^*(o,t)\} = \sum_u \mathrm{E}\{N^*(o,t) \mid \delta(u,t) = 1)\ \Pr\{\delta(u,t)\}.$$

According to the probability model Π, $\Pr\{\delta(u,t)\} = 1/U_o$. Then, given L.3 and D.7, we have

$$\mathrm{E}\{N^*(o,t)\} = (1/U_o) \sum_u c^*(u) = I^*_o.$$

Reference to D.7 completes the regularity chain that supports the theorem.

When we relate opacity to the expected duration of a cascade, we will need an expression that specifies the probability that an induced violation is foreseen. As with the assumptions in the probability model Π, we want to simplify as much as possible. Therefore, we introduce an auxiliary assumption that holds that each induced violation is equally likely to be foreseen and that this probability equals the fraction of foreseen to total induced violations. Let $v^*_u(u',t)$ be a random variable that equals one if the pair $\langle u',t' \rangle$ is one of the induced violations in the cascade $\mathrm{K}(u,t)$ and the occurrence of this induced violation is foreseen from the vantage point of u, the initiator of the cascade and that equals zero if it is not foreseen.

Postulate 1 (P.1). The expected duration of an unforeseen violation normally exceeds that of a foreseen one:

$$\mathfrak{N}o,u,u',u'',t[\mathrm{O}(o,p) \wedge \mathrm{U}(u,o) \wedge \mathrm{U}(u',o) \wedge \mathrm{U}(u'',o) \wedge (\delta(u,t) = 1) \rightarrow \mathrm{E}\{d(u',t' \mid v^*_u(u',t') = 1)\} < \mathrm{E}\{\mathrm{d}(u'',t'' \mid v^*_u(u'',t'') = 0)\}].$$

Finally, we need to define the probability that an actual violation is foreseen.

Auxiliary Assumption 1 (A.1). The probability that an initiating unit in a cascade foresees an induced violation does not vary over units or over time; it equals the ratio of obscured intricacy to actual intricacy:

$$\mathfrak{N}o,u,u',u'',t,t'[\mathrm{O}(o,p) \wedge \mathrm{U}(u,o) \wedge \mathrm{U}(u',o) \wedge \mathrm{U}(u'',o) \wedge (\Delta(o,t) = 1) \rightarrow \Pr\{v^*_u(u',t') = 1\} = \Pr\{v^*_u(u'',t'') = 0\} = I^*_o / I_o].$$

CULTURAL OPPOSITION: ASPERITY

Our second major extension to the model of cascading organizational changes takes seriously the possibility of cultural opposition to an architectural change. By cultural opposition, we mean that the proposed architecture violates the cultural code of the organization, causing normative reactions against the change. These reactions could emanate from a variety of sources, including local tradition, professional norms, identity and form constraints, and national culture and social structure (Crozier, 1964). From our point of view, the important observations are that (1) the strength of cultural opposition can be difficult to anticipate, (2) normative opposition generates turmoil, and (3) turmoil likely lengthens the reorganization period.

We want to consider situations in which units experience induced violations of architectural codes and take actions to modify the local architecture to come into conformity with the newly imposed external constraints. We do not attempt to specify exactly when the cultural code violation occurs, other than to assume that it occurs after the induced violation of architectural code. We distinguish cases in which the local modifications of the architecture conform to the local culture from those in which it does not. Let the random variable $y(u,t)$ equal one if unit u experiences cultural opposition during an attempt to resolve an induced violation of architectural code at time t and equal zero otherwise. When such conflict occurs, the new architecture becomes morally suspect in the eyes of a unit's members. We believe that this was the case in the examples mentioned above: CEO Amelio's attempt to centralize Apple Computer and the top management of Ben & Jerry's Ice Cream's attempt to expand the internal salary ratio. Note that there might be nothing operationally wrong with the new code; the architectural change might very well prove beneficial except for the cultural reaction.

In analyzing cultural opposition, we build on the concept of asperity, the fraction of officially allowed architectures that the culture rules out. Because a rigorous culture rules out many architectures, efforts to modify architecture more likely violate cultural rules in organizations with greater asperity. We formalize this idea as follows.

Postulate 2 (P.2). The probability that an induced architectural change will trigger cultural opposition to local architectural change is proportional to the asperity of a unit's culture:

$$\mathfrak{R}\, o,u,u',t[\mathrm{O}(o,p) \wedge \mathrm{U}(u,o) \wedge (v(u,u',t) = 1) \rightarrow \Pr\{y(u,t)\} = a(u,t)].$$

Cultural opposition resists ready resolution for at least three reasons. First, cultural violations often produce intense moral reactions that cause individuals and groups to fight harder and longer. Second, those outside a unit might not perceive that the opposition is cultural because even cultural resistance often concentrates on technical matters. So resolution efforts might focus mistakenly on non-cultural issues perceived to be the source of opposition. Third, organizational culture provides a potential basis for very broad resistance. Thus the proponents of the changes cannot count on many members of the organization to use their specialized knowledge to facilitate the many local adjustments required to complete the change. Such changes require more direct managerial control. Given limits on managerial time and effort, reliance on managerial control puts sharp limits on the pace of adjustment. Accordingly, we postulate that cultural resistance slows processes of change.

Postulate 3 (P.3). It generally takes longer to eliminate an induced violation of an architectural code if cultural codes get violated:

$$\mathfrak{N} o,u,u',t,t'[\mathrm{O}(o,p) \wedge \mathrm{U}(u,o) \wedge \mathrm{U}(u',o) \wedge (y(u,t) = 1) \wedge (y(u',t') = 0) \rightarrow \mathrm{E}\{d(u,t)\} > \mathrm{E}\{d(u',t')\}].$$

We now want to aggregate to the organizational level under the simplifying (auxiliary) assumption that organizations have characteristic levels of asperity that do not vary among units or over time. This assumption is clearly justified in at least some cases by the extensive literature on organizational culture positing the existence of unitary cultural characteristics.

Auxiliary Assumption 2 (A.2). Each organization has a characteristic (common) level of asperity that applies to all units:

$$\mathfrak{N} o \exists A_o \forall u,u',t,t'[\mathrm{O}(o,p) \wedge \mathrm{U}(u,o) \wedge \mathrm{U}(u',o) \rightarrow (a(u,t) = a(u',t') = A_o)].$$

OPACITY, ASPERITY, AND THE DURATION OF REORGANIZATION

We now state explicit assumptions relating opacity and asperity to the time it takes for an induced architectural violation in a unit to be resolved. Stage 1 of the theory used an auxiliary assumption that holds that the expected duration of an induced violation does not vary over units or over time for an organization. This common duration was labeled as τ_o and called viscosity (sluggishness). At this point, we take advantage of nonmonotonic logic to override the earlier assumption with new ones that apply to situations of greater specificity. The original assumption continues to hold and is not overridden if *all* that we know is that an organization belongs to a given population. If we know more than this, specifically, if we have information about opacity and asperity, then the original default is replaced by the more-specific auxiliary assumption stated below.

Our conceptual framework highlights four conditions: an induced architectural violation for a unit can be either foreseen or not, and it can either trigger cultural resistance to change or not. One approach to modeling these conditions assumes that the effects of opacity and asperity are additive. Then we need only three organization-specific parameters: a baseline duration, in which neither complication occurs, an effect of opacity, and an effect of asperity. Then the four possibilities get one, two, or three of the parameters. This kind of model is appropriate when the effect of opacity does not depend on the level of asperity. Because this kind of independence might not hold in cases of interest, we impose a weaker four-parameter specification. The fourth parameter (labeled as $\tilde{\rho}_o$ below) can be either positive or negative, but in this case its value is constrained, as we explain below. A negative value means that the combination of complications can be dealt with more quickly than when the complications arise singly. A positive value means that the combination lengthens durations more than would be expected from knowledge of the separate effects. We think that positive values of this parameter are likely in real applications.

Auxiliary Assumption 3 (A.3). The expected duration of a reorganization in a unit depends on (1) whether the induced violation is foreseen and (2) whether the response to the induced violation encounters culturally based opposition. In combinations of these two conditions, the expected duration is constant over pairs of units and over time points. Thus, there are four organization-specific (constant) characteristic durations.

Assume that four units in an organization experience induced violations.

$$\begin{gathered}\mathfrak{A} o \exists \rho_o, \tilde{\rho}_o, \theta_o, \tilde{\theta}_o \forall u_1, u_2, u_3, u_4, t_1, t_2, t_3, t_4 [\mathrm{O}(o,p) \wedge \\ \mathrm{U}(u_1,t_1) \wedge \mathrm{U}(u_2,t_2) \wedge \mathrm{U}(u_3,t_3) \wedge \mathrm{U}(\mathrm{u}_4,\mathrm{t}_4) \wedge \\ (v^*_u(u_1,t_1) = 1) \wedge (y(u_1,t_1 = 0) \rightarrow \mathrm{E}\{d(u_1,t_1) \mid \delta(u,t) = 1\} = \theta_o) \wedge \\ (v^*_u(u_2,t_2) = 1) \wedge (y(u_2,t_2 = 1) \rightarrow \mathrm{E}\{d(u_2,t_2) \mid \delta(u,t) = 1\} = \theta_o + \rho_o) \wedge \\ (v^*_u(u_3,t_3) = 0) \wedge (y(u_3,t_3 = 0) \rightarrow \mathrm{E}\{d(u_3,t_3) \mid \delta(u,t) = 1\} = \theta_o + \tilde{\theta}_o) \wedge \\ (v^*_u(u_4,t_4) = 0) \wedge (y(u_4,t_4 = 1) \rightarrow \mathrm{E}\{d(u_4,t_4) \mid \delta(u,t) = 1\} = \theta_o + \tilde{\theta}_o + \rho_o + \tilde{\rho}_o)].\end{gathered}$$

Because a duration cannot be negative, $\theta_o \geq 0$. P.1 states that unforeseen induced violations take longer to resolve than foreseen ones, which implies that $\rho_o > 0$ (otherwise P.1 would be violated). Similarly, P.3 states that induced violations whose attempted resolution meets cultural resistance last longer than those that do not. This implies that $(\tilde{\theta}_o > 0) \wedge (-\rho_o < \tilde{\rho}_o)$. We record these constraints in the following lemma.

Lemma 5 (L.5).

$$\mathfrak{P} p,o[\mathrm{O}(o,p) \rightarrow (\rho_o > 0) \wedge (\tilde{\theta}_o > 0) \wedge (-\rho_o < \tilde{\rho}_o)].$$

REORGANIZATION PERIODS

The central tenets of our argument hold that the expected duration of a reorganization in a random cascade increases with opacity and asperity. In the case of opacity, this is because the agents cannot know a priori all of the adjustments required to eliminate architectural code violations in an opaque organization. As a result, not all changes can be undertaken in parallel. Only when a cascade of adjustments within one part of the organization has played itself out does a downstream unit have a clear understanding of the new constraints.

A potentially promising way of thinking formally about the length of the period of reorganization in these cases is to view the subject as a queuing problem. Queuing theory provides a framework for analyzing the behavior of queues including waiting times with uncertain (stochastic) arrivals. In the usual queuing problem setup, the main modeling choices concern (1) the rate and stochastic form of arrivals to the system; (2) the rate at which arrivals get served and the corresponding waiting times; and (3) the number of service agents available. There are also endless potential complications such as queue discipline and waiting behavior.

Consider the extreme case of an organization for which all adjustments must be done individually and in sequence, but the recognition of what adjustments need to be made is partial and evolves randomly over time. The required individual adjustments might be seen as arrivals to a queuing system. Because the adjustments can only be solved one at a time, the system can be seen as possessing a single service agent. With further specification of the exact form of the arrival process and the agent's speed, a variety of useful analytical results can then be obtained (see, e.g., Gross and Hart, 1985). For example, one could derive the average waiting time from recognition to completion of a required adjustment, as well as behavior of the whole system in terms of its busy and idle periods, representing the time spent in reorganization. Further analyses could relax the assumptions of this simple illustration. At this point in the development of the theory, we concentrate on the simpler implications of opacity and asperity.

Hence, induced violations will arise at random. A unit might be well on its way to resolving an induced violation when another violation gets induced over another path in the pattern of architectural dominance relations. Hannan, Pólos, and Carroll (2003a) argued that such a situation takes at least twice as long to resolve as would an otherwise similar uninterrupted spell of resolution. For example, when Hewlett Packard spun off what became known as Agilent Technologies, the corporate headquarters of the new organization consolidated and centralized certain shared functions such as information technology that had been fully decentralized in the parent company. For most functions, the consolidation process apparently took months to achieve. The implications of these changes for the various businesses will take much longer to work out, because many units are somewhat opaque to the central administration. The company faces the real danger that the tightening of local resources might diminish innovation in the long run, since the units historically had almost full autonomy and did not need to ask for resources to attempt experimental projects as they might now.

The time required to resolve an inconsistency generally grows with asperity. We argued above that this is because the reactions are more intense, the opposition might not be correctly identified as cultural, and the broad basis of opposition might require more managerial attention in fine-tuning the structure in adjustment. Of course, holding constant the number of induced violations, the total time in reorganization will increase with the strength of the cultural opposition. Coupling these ideas with the earlier argument about the intricacy of organizational design yields the following theorem.

Theorem 1 (T.1). The expected total time spent in reorganization by an organization's units during a random cascade increases monotonically with intricacy, I_o, opacity, P_o, and asperity, A_o.

$$\mathfrak{P} o,t[\mathrm{O}(o,p) \wedge (\Delta(o,t) = 1) \rightarrow \mathrm{E}\{D(o,t)\} = \theta_o I_o + \tilde{\theta}_o P_o + (\rho_o I_o + \tilde{\rho}_o P_o)A_o].$$

Proof. By definition, $\mathrm{E}\{D(o,t)\} = \mathrm{E}\{\Sigma_{\langle u',t'\rangle \in \mathbf{K}(u,t)} d(u',t')\}$. The summation ranges over all of the induced violations in the cascade. According to L.2, the number of induced violations in a cascade with random origin is given by I_o. Therefore,

$$\mathrm{E}\{D(o,t)\} = I_o\, \mathrm{E}\{d(u',t')\}.$$

According to the law of total probability (and given the premise that a change did occur at t),

$$\begin{gathered}
\mathrm{E}\{d(u',t')\} = \\
\mathrm{E}\{d(u',t' \mid v^*_u(u',t') = 1) \wedge (\mathrm{y}(u',t') = 0)\}\ \Pr\{v^*_u(u',t'') = 1) \wedge \\
(y(u',t') = 0)\} + \mathrm{E}\{d(u',t' \mid v^*_u(u',t') = 1) \wedge (\mathrm{y}(u',t') = 1)\}\ \Pr\{v^*_u(u',t') = 1) \wedge \\
(y(u',t') = 1)\} + \mathrm{E}\{d(u',t' \mid v^*_u(u',t') = 0) \wedge (\mathrm{y}(u',t') = 0)\}\ \Pr\{v^*_u(u',t') = 0) \wedge \\
(y(u',t') = 0)\} + \mathrm{E}\{d(u',t' \mid v^*_u(u',t') = 0) \wedge (\mathrm{y}(u',t') = 1)\}\ \Pr\{v^*_u(u',t') = 0) \wedge \\
(y(u',t') = 1)\}.
\end{gathered}$$

Given Π, A.1 and A.2, the foregoing can be expressed as

$$\begin{gathered}
\mathrm{E}\{d(u',t')\} = \\
\mathrm{E}\{(d(u',t' \mid v^*_u(u',t') = 1) \wedge \\
(\mathrm{y}(u',t') = 0)\}\ (I^*_o / I_o)(1 - A_o) + \mathrm{E}\{(d(u',t' \mid v^*_u(u',t') = 1) \wedge \\
(\mathrm{y}(u',t') = 1)\}\ (I^*_o / I_o)A_o + \mathrm{E}\{(d(u',t' \mid v^*_u(u',t') = 0) \wedge \\
(\mathrm{y}(u',t') = 0)\}\ ((I_o - I^*_o)/ I_o)(1 - A_o) + \mathrm{E}\{(d(u',t' \mid v^*_u(u',t') = 0) \wedge \\
(\mathrm{y}(u',t') = 1)\}\ ((I_o - I^*_o)/ I_o)A_o.
\end{gathered}$$

Using the expected durations given by A.3, we have:

$$\begin{gathered}
\mathrm{E}\{\mathrm{D}(\mathrm{o},\mathrm{t})\} = I_o(\theta_o(I^*_o / I_o)(1 - A_o) + (\theta_o + \rho_o)(I^*_o / I_o)A_o + \\
(\theta_o + \tilde{\theta}_o)((I_o - I^*_o) / I_o)(1 - A_o) + (\theta_o + \tilde{\theta}_o + \rho_o + \tilde{\rho}_o)(I_o - I^*_o) / I_o)A_o).
\end{gathered}$$

Using the definition of organizational opacity in D.7 (and rearranging terms) concludes the proof.

Consider the relevant cases in light of this theorem. The expected total time reorganizing during a random cascade equals the sum of three terms. The first, $\theta_o I_o$, gives the expected duration for an organization that is fully transparent, $P_o = 0$, and whose organizational culture admits all possible architectures, $A_o = 0$. Therefore, this result shows that the result for Stage 1 of the theory persists as a special case of transparency and no cultural restraint in the new formulation.

Suppose that transparency is incomplete, $P_o > 0$, but cultural restraint is lacking, $A_o = 0$. Then the expected total time reorganizing equals $\theta I_o + \tilde{\theta}_o P_o$. Given the definition of opacity and L.3, this result says that the expected total time increases monotonically with opacity in the absence of cultural restraint.

Finally, bring asperity into the picture. The final term to be considered, $(\rho_o I_o + \tilde{\rho}_o P_o)A_o$, shows that asperity interacts with intricacy and opacity in affecting the expected time reorganizing. The coefficient of the intricacy-asperity effect, P_o, gives the expected duration of a foreseen induced violation that involves a cultural violation. The coefficient of the opacity-asperity effect, $\tilde{\rho}_o$, A_o gives the expected duration of an unforeseen induced violation that involves a cultural violation. A.3 and L.5 say that all of the parameters, other than $\tilde{\rho}_o$, are positive. Even if $\tilde{\rho}_o < 0$, the combination of the last constraint expressed in L.5 and the definition of P_o imposes that the overall effect of P_o and A_o be positive. Therefore, the expected total time reorganizing increases monotonically with intricacy, opacity, and asperity. It is nonetheless important to remember that these three factors might not vary independently. In particular, intricacy and opacity are definitionally dependent: a non-intricate organization cannot be opaque.

MISSED OPPORTUNITIES

We focus on periods of reorganization because devoting attention, time, and energy to reorganization (adjusting codes to eliminate induced violations) diverts members of an organization from the tasks that generate revenues. Therefore, any lengthy change process generally entails substantial opportunity costs. Opportunities are missed because managerial attention focuses on managing the change, production gets disrupted, relations with customers are left unattended as responsibilities are reallocated, and so forth. Each of these problems becomes more serious as a reorganization lengthens. During a reorganization period, considerable attention is paid to the fate of the new architecture. The units with changed architectural codes generally face scrutiny, and nonconformity to the newly added code is noted. In other words, unlike normal functioning, in which managerial attention to architectural conformity is partial and episodic, violations of newly added code generally get noticed during a reorganization period. The resulting diversion of attention causes opportunities to be missed.

We specify the consequences of missed opportunities in terms of the growth rate of resources. Let $R(o,t)$ denote the random variable that records the level of organization o's resources at time t. Organizations lose resources during periods of reorganization, both because reorganization is costly and because directing resources and attention away from production dampens resource acquisition.

According to the probability model Π, opportunities arrive at the same constant rate, Ξ, for each organization in a population. An organization's architecture determines what part of the organization can see a given opportunity; at the extreme, in an organization with only one unit, all of the relevant opportunities arrive at the same unit. Π includes the assumption that units within an organization are equally likely to experience the opportunity. Therefore, the arrival rate of opportunities to a unit in an organization with U_o units is $\xi \equiv \Xi \,/\, U_o$.

We want to focus on the expected difference in the number of opportunities missed over an interval for a reorganizing unit and an otherwise identical not-reorganizing one. Let the random variable $m(u,w,w')$ give the number of opportunities missed by unit u during the arbitrary interval $[w,w')$ and let the organization-level counterpart by $M(o,w,w') \equiv \Sigma_{u:\mathrm{U}(u,o)} m(u,w,w')$. We want to characterize the expected value of $m(u,w,w')$.

We begin by comparing two organizations, one that is reorganizing over the whole period being considered and one that is not reorganizing at any time during the period.

Lemma 6 (L.6). (L.5 in Hannan, Pólos, and Carroll, 2003a)

$$
\begin{gathered}
\mathfrak{P}\, o,u,u',t_1,t_2,t_3,t_4[\mathrm{O}(o,p) \wedge \\
\mathrm{U}(u,o) \wedge \mathrm{U}(u',o) \wedge \forall w,w'[(t_1 \le w < t_2) \wedge \\
(t_3 \le w' < t_4) \rightarrow (\mathrm{RE}(u,w) \wedge \\
\neg\mathrm{RE}(u',w')] \rightarrow (\mathrm{E}\{m(u,t_1,t_2)\} = \xi_o(\eta + \tilde{\eta})(t_2 - t_1)) \wedge \\
(\mathrm{E}\{m(u',t_3,t_4)\} = \xi_o\eta(t_4 - t_3))].
\end{gathered}
$$

Next we want to extend the result to apply to a full cascade of changes for an organization.

Theorem 2 (T.2). The expected number of opportunities missed during a full cascade is the sum of (1) the baseline expected number that would be missed over an interval had no reorganization taken place and (2) an excess expected number that increases monotonically with the product of the organization's intricacy, I_o, opacity, P_o, and asperity, A_o.

$$
\begin{gathered}
\mathfrak{P}\, o,u,t[\mathrm{O}(o,p) \wedge \mathrm{U}(u,o) \wedge (\Delta(o,t) = 1) \rightarrow \\
\mathrm{E}\{M(o,t,t + S(o,t))\} = \Xi\eta S(o,t) + \xi_o\theta(\eta_o I_o + \tilde{\eta}_o P_o + (\rho_o I_o + \tilde{\rho}_o P_o)A_o)].
\end{gathered}
$$

Proof. The most-specific regularity chain combines the most specific regularity chains supporting L.6 and T.1. According to L.6, each unit contributes $\xi_o\eta_o S(o,t)$ expected missed opportunities, the baseline that holds whether or not a unit is in reorganization. This part of the process therefore contributes $\xi_o\eta \mathrm{S}(o,t)U_o = \Xi\eta_o S(o,t)$ expected missed opportunities. Next consider the additional expected missed opportunities due to reorganization. The expected total time spent reorganizing in a complete cascade equals $\theta(\eta_o I_o + \tilde{\eta}_o P_o) + (\rho_o I_o + \tilde{\rho}_o P_o)A_o$ according to T.1. The expected number of excess opportunities missed during reorganization is this expected time multiplied by ζ_o.

We do not assume any knowledge of the function that relates missed opportunities to the growth in resources. Instead we argue for a weaker monotonicity relationship, which affects what follows: instead of getting precise results about expectations, we get monotonicity statements relating intricacy, opacity, and asperity to resource growth and mortality hazards.

Postulate 4 (P.4). Consider a pair of organizations in the same population with equal stocks of resources at the start of a time interval. If one misses more opportunities over the interval than the other, then its expected growth in resources is lower. Otherwise, the expected growth in resources for the two is the same. (The equality condition follows from the background assumption that the trichotomy relation holds, that if it is not the case that $a > b$ or $a < b$, then $a = b$.)

$$\mathfrak{R}\, o,o',s,s',w,w'[\mathrm{O}(o,p) \wedge \mathrm{O}(o',p) \wedge \\ (R(o,s) = R(o',s')) \wedge \\ (M(o,s,w) > M(o',s',w')) \rightarrow E\{R(o,w)\} < E\{R(o',w')\}].$$

We can use the foregoing argument to compare what happens to two organizations that experience cascades of change. A subtle issue needs attention. The two cascades might differ in temporal scope, $S(o,t)$. We want to compare experiences over the time span of the longer cascade, so that we get the full scope of both cascades. Things get very complicated if we allow the possibility that the organization with the shorter cascade starts another cascade within the period of comparison. So we want to restrict the comparison to the case in which no subsequent initiations of cascades occur within the period of comparison.

Notation. To avoid repeating a very complicated expression in a series of lemmas and theorems, we introduce some notational shorthand.

$$Z = \max(S(o,t), S(o',t')).$$

The formula Ψ stands for the following formula: the entities being compared, o and o', are organizations in the same population with equal resources, experience architectural changes at times t and t', respectively, and neither experiences another (uninduced) architectural change until the end of the longer of the two cascades of change.

$$\Psi \leftrightarrow \mathrm{O}(o,p) \wedge \mathrm{O}(o',p) \wedge (\Delta(o,t) = \Delta(o',t') = 1) \wedge \\ (R(o,t) = R(o',t')) \wedge \forall s,s'[(t < s \leq (t + Z)) \wedge$$

$$(t' < s' \leq (t' + Z)) \rightarrow \Delta(o,s) = \Delta(o',s') = 0].$$

The key implication of this argument can be captured by the following lemma.

Lemma 7 (L.7). An organization's expected rate of growth in resources during a cascade with a random origin decreases with the total time reorganizing, $D(o,t)$ (L.6 in Hannan, Pólos, and Carroll, 2003a):

$$\mathfrak{N}\, o,o',t,t'[\Psi \wedge (D(o,t) > D(o',t')) \rightarrow E\{R(o,t + Z)\} < \mathrm{E}\{R(o',t' + Z)\}].$$

CHANGE AND ORGANIZATIONAL MORTALITY

Hannan and Freeman's (1984) theory of structural inertia assumed that the hazard of mortality rises monotonically with the duration of the period of reorganization. We now show that this premise follows as a theorem in the new theory. Let $\mu(o,t)$ denote the mortality hazard for organization o at time t. A hazard equals the ratio of the probability density of the length of lifetimes in a population to the survivor function. In other words, the hazard (at a particular time) provides a local characterization of the probability distribution of length of lifetimes. This means that we can formulate propositions and theorems about hazards directly in terms of the hazard function, rather than in terms of expectations. (If we wanted to formulate this part of the argument in terms of expected values, then we would consider expected lifetimes.) Nearly all treatments of the relationship of size and resources with mortality assume that organizations with access to greater resources can better withstand life-threatening environmental shocks (Carroll and Hannan, 2000).

Postulate 5 (P.5). An organization's hazard of mortality declines monotonically with its level of resources (P. 3 in Hannan, Pólos, and Carroll, 2003a):

$$\mathfrak{N}\, o,o',s,s'[\mathrm{O}(o,p) \wedge \mathrm{O}(o',p) \wedge (R(o,s) > R(o',s')) \rightarrow \mu(o,s) < \mu(o',s')].$$

The argument in the preceding sections identifies structural and cultural features that lengthen expected durations of cascades. If stocks of resources fall monotonically during cascades, then the factors that lengthen cascades make change more risky. We now provide a formal statement of these implications for pairs of organizations that experience random architectural changes and differ in the structural and cultural factors that affect expected cascade length. As noted above, these theorems give monotonicity relations. Each considers one of the structural/cultural factors in isolation from the others. This is because the theory contains regularity chains that warrant inferences about such relations.

Theorem 3 (T.3). The increase in the hazard of mortality due to an architectural change grows monotonically with intricacy.

$$\mathfrak{P}o,o',t,t'[\Psi \wedge (I_o > I_{o'}) \rightarrow \int_{t}^{t+Z} \mu(o,v)dv > \int_{t'}^{t'+Z} \mu(o',v)dv].$$

Proof. The most-specific regularity chain begins with the intension of $I_o > I_{o'}$. T.1 ties this condition to the expected length of a reorganization period. In the next step, we assume that the facts agree with the expectations: that the actual duration for the more intricate organization exceeds that of the less intricate one. It follows that there is a period during which both organizations are in reorganization. Under the assumption that they started with the same resource levels, according to P.4, the expected resource levels of the pair are the same when the reorganization for the less intricate organization ends. From that time point until the end of reorganization for the more intricate organization, the expected resource levels fall for the latter but not for the former, according to this postulate. Again if the facts agree with the difference in expected values, then the mortality hazard for the more intricate organization exceeds that of the less intricate one over this period, by P.5. Hence, the regularity chain states that the hazards of the two do not differ over the initial part of their common period of reorganization but do differ afterwards, until the reorganization for the more intricate organization ends. The regularity chain supports the claim of the theory. The available postulates and auxiliary assumptions do not support any countervailing regularity chain.

Theorem 4 (T.4). The increase in the hazard of mortality due to an architectural change grows monotonically with opacity.

$$\mathfrak{P}o,o',t,t'[\Psi \wedge (P_o > P_{o'}) \rightarrow \int_{t}^{t+Z} \mu(o,v)dv > \int_{t'}^{t'+Z} \mu(o',v)dv].$$

Proof. The proof follows the lines of the proof of T.3.

Theorem 5 (T.5). The increase in the hazard of mortality due to an architectural change grows monotonically with asperity.

$$\mathfrak{P}o,o',t,t'[\Psi \wedge (A_o > A_{o'}) \rightarrow \int_{t}^{t+Z} \mu(o,v)dv > \int_{t'}^{t'+Z} \mu(o',v)dv].$$

Proof. The proof follows the lines of the proof of T.3.

Now we can see another advantage of applying the nonmonotonic logic in formalizing organization theory. Theorems 3, 4, and 5 state that intricacy, opacity, and asperity increase the hazard due to change. Suppose that we encounter a class of simple (low in intricacy) but opaque (or simple and high in asperity) organizations. What do we predict? Two arguments that do not clearly differ in specificity point in opposite directions, a case that logicians call a Nixon diamond (see Pólos and Hannan, 2002, for details). The lack of an established specificity difference leads us, according to the methodological principle embedded in the logic, to refrain from making a prediction (an inference, more generally). More information is needed, perhaps concerning the exact functional form of the relationships involved. In any event, given the present state of knowledge (as reflected in the premises of the theory), no prediction can be offered.

If this result seems undesirable, consider the main alternative. Suppose that we had constructed the theory in classical first-order logic. Then the three universally quantified theorems (stated in terms of comparisons of strictly monotonic functions for intricacy, opacity, and asperity) would imply that there could not be any pairs of organizations such that one is more intricate and the other has greater opacity or that one has greater opacity and the other has higher asperity and so forth (see Péli, Pólos, and Hannan, 2000: 201). Consider a simple example. Suppose we have three functions, $p(\cdot)$, $q(\cdot)$, and $r(\cdot)$, and the following postulates:

$$\forall x,y[p(x) > p(y) \rightarrow q(x) > q(y)];$$

$$\forall x,y[r(x) > r(y) \rightarrow q(x) > q(y)].$$

A logically equivalent syntactic variant of the first postulate is

$$\forall x,y[p(y) > p(x) \rightarrow q(y) > q(x)].$$

From the second postulate, by contraposition, we have:

$$\forall x,y[\neg(q(x) > q(y)) \rightarrow \neg(r(x) > r(y))].$$

The antecedent of this formula can be satisfied if $q(y) > q(x)$. Therefore, by the chain rule one can derive

$$\forall x,y[p(y) > p(x) \rightarrow \neg(r(x) > r(y))].$$

This form of deduction is not warranted in the nonmonotonic logic, because this logic does not support the contraposition operation (on the grounds that reasoning about causal stories does not warrant it).

We conclude that a first-order version of the theory is too strong for the knowledge base on which it is built. It implies surprising theorems: intricacy is monotonically related to opacity and also to asperity, and so forth. Do we want such theorems? Clearly not. We think that these conclusions might easily turn out to be false empirically. Moreover, they do not appear to reflect the intuitions that motivate the theory. Having such theorems as a permanent part of the theory would impose undesirable constraints on further development of the theory. So reliance on a first-order formulation would have us painting ourselves into a corner from the perspective of future development of the theory. None of these undesirable consequences arise in the (weaker) nonmonotonic formulation.

THE BARINGS BANK COLLAPSE—AN INTERPRETATION

The spectacular collapse of Barings Brothers, a venerable British bank, in February 1995 is widely blamed on a single "rogue" trader named Nicholas Leeson, who worked in Singapore. While there is no doubt that Leeson's activities directly generated the losses that caused the bank to fail, the conditions under which such activities could occur and persist undetected is plausibly explainable by our theory: the same activities attempted in a different organization or in Barings at a different time would likely have been caught early on and stopped long before they led to huge losses, let alone took down the entire bank.

Nicholas Leeson worked for Baring Futures Singapore (BFS), which was an entity operated by Baring Securities Ltd. (BSL). BSL originated in 1984 as the subsidiary Baring Far East Securities (it was renamed in 1986) when the executives of Baring Brothers acquired a small stockbrokering firm named Henderson Crosthwaite in anticipation of the United Kingdom's deregulation of financial markets in 1986, the so-called Big Bang. Before the Big Bang, British merchant banks could issue securities, but they could not trade or distribute them. Barings had been a major player in the old system—since 1762 it had been run as a family-run merchant bank with an enviable client list including royalty from several countries and, more recently, eminent international firms.

The Big Bang promised to modernize British financial markets. Like Barings, most banks made plans to enter the newly deregulated areas, either on their own through expansion or acquisition or through the establishment of alliances with brokerage partners. Other banks typically set up more integrated structures for taking on the newly allowed brokering-trading activities. The original arrangement at Barings was notable for the independence it offered BSL:

> The new subsidiary was to be completely autonomous, with separate management and office. In order to give the employees a direct interest in the annual results of the firm, it was to retain 50 percent of its profits before tax, after a capital charge, for distribution among its staff, with the other 50 percent going to Baring Brothers & Co., its parent. (Rawnsley, 1995: 39)

The autonomy offered in the Barings structure proved to be both a blessing and a curse. In a short time, Barings was able to move quickly into the new activities and to avoid the types of organizational and cultural conflicts hampering banks that attempted to build integrated structures. Looking back on the first decade of operation, BSL's founder concluded:

> We had an environment in which we could build the business, in that we were totally independent of Baring Brothers, so we had no daily cultural battle to fight, which a lot of other firms had after the Big Bang when they tried to weld a whole lot of people who were not suited to working together. The merchant banking culture and the brokering culture are very different at the end of the day. (quoted in Rawnsley, 1995: 40)

It helped, too, that much of the new unit was located halfway around the world, based in Japan. Given such autonomy, BSL was able to trade at prices Barings would have found horrifying, to develop a strong customer service orientation, and to discover and reap newly developing markets such as that of Japanese warrants. The result was a highly profitable, if high-flying and, occasionally, chaotic entity. Both returns and firm growth were dramatic in the early years. Before the market crash of October 1987, "Securities were providing over two-thirds of the bank's profits" (Rawnsley, 1995: 66).

By the 1990s, the bubble clearly had deflated. Envy within the bank and elsewhere developed when freewheeling BSL executives were publicly reported to have some of the highest compensation in the U.K. (in 1986, the highest paid man at BSL did in fact hold the top position). The market downturn in 1990, coupled with the associated major losses in warrants, made BSL a less attractive subunit to Barings officials. BSL argued for further aggressive growth, including movement into proprietary trading, which required additional infusions of capital from London. Taking on added risk while faced with a lack of information and control did not seem a good proposition to the Barings people, many of whom had been waiting for their chance to reel BSL back in. Even the BSL chief admitted that the operation had grown larger than the existing organizational and management controls could handle. At this juncture, both Barings and BSL wanted to formalize the organization, moving it away from the entrepreneurial structure that had been so conducive to growth. But BSL wanted to retain autonomy, while Barings insisted that the two entities be integrated. A comprehensive study of BSL was initiated after some unexpected losses coincided with the awarding of large executive bonuses. The subsequent report revealed "the absence of controls; there was no business plan or strategy, no effective control system or budgets, no management, offices had been opened all over the place . . . [with the chief executive officer] trying to micro-manage it" (quoted in Rawnsley, 1995: 95).

Two actions were taken in response to these events. First, the BSL unit was "solo consolidated" within the Barings Bank. This action mainly addressed regulatory issues. Solo consolidation allowed the bank to treat BSL as a branch of the bank for regulatory reporting purposes. This meant that loans made to the BSL from the parent bank "would not be subject to the large exposures, reporting requirements and constraints" (San and Kuang, 1995: 97). In reporting to the Bank of England, the two entities would file only a single return. Among the restrictions for solo consolidation are that the unit be wholly funded by its parent, that it be 75 percent owned by the parent, that the parent maintain effective control, possess sufficient capital, and that there should be no obstacles to transfers such as overseas exchange limits. Second, a major restructuring plan was adopted in fall 1992, with the goal of merging the bank and BSL into an integrated investment bank. The plan would reduce costs and staff by some 15–20 percent, closing some offices and selling off others. Warrant trading was to be shut down. The savings was projected at 20 million pounds annually. Barings was then to invest another 45 million pounds into the business. Within three years, BSL was to be fully merged into the newly integrated investment bank.

The author of the Barings study assumed the position of chief operating officer (COO) and began restructuring activities. The new environment has been described as a "political hotbed." Within six months, the founding director of BSL resigned, along with other key figures. Senior management was then restructured, with the COO appointed as chief executive officer (CEO) and other Barings people forming a new top management team at BSL. The new CEO himself was not familiar to the BSL staff, personally or professionally.

The effort to integrate the two units involved the establishment, in 1993, of an umbrella organization known as Baring Investment Bank, composed of four major groups. The two largest groups were the Bank Group and the Equity Brokering and Trading Group. The main idea was to "coordinate product activities on a global basis along with the decentralized authority" (Rawnsley, 1995: 128). Management authority was to be delegated locally to four regional centers in Tokyo, Hong Kong, Singapore, and New York. Profits were to be the responsibility of product groups, while the local offices were to provide infrastructure and facilitate information sharing. A matrix reporting system was put in place to provide stronger controls and more accountability.

The expected difficulties of this integration seem to have been anticipated in the announcement of a three-year window for implementation. Still, it is doubtful that the confusion abounding during the reorganization period had been anticipated. The cultures proved to be hostile to each other, and many key employees resented integration. The old support systems (accountancy, treasury, and risk) were found to be outdated and in need of updating before the new controls could work. At the same time, the firm was pushing ahead as fast as possible to develop new business, often outrunning its support structure. According to Rawnsley (1995: 133),

> . . . the new reporting lines and responsibilities were not perceived to be clear by many, nor were they fully understood. It appears that no complete organizational chart was prepared and disseminated when the organizational structure was being formulated . . . this situation was exacerbated in Baring Securities' Asian offices where a series of personnel changes took place.

It was in this context of urgency mired in confusion that Leeson worked. He understood the need to show profits, and he also realized that the formal and informal controls left much room for maneuvering when his returns were not up to par. Among other things, he was in charge of his own trading settlements, something simply not allowed in the securities business. It was also unclear who was supposed to supervise him, the local boss or the product group supervisor. As long as he showed profits on paper, it seemed that no one cared. Moreover, many were willing to run interference for him, allowing him to get more capital and to avoid the scrutiny of London, in part because the "profits" he generated contributed to their bonuses. So, when the market turned against him, Leeson was able to make larger and larger bets using the firm's capital. When the bets challenged the bank's entire capital structure, they finally were detected. But, of course, it was then too late—Barings was bankrupt.

Without denying Leeson's guilt, most subsequent assessments of the meltdown attribute the failure to the reorganization and its attendant problems. For instance, Rawnsley (1995: 184) concluded that "during this state of flux, the organization was in a position of considerable vulnerability; adequate controls were not in place quickly enough for management to detect the trading losses." She quotes George MacLean, head of the Bank Group: "I believe the seeds of this [collapse] were sown when we went into Baring Securities Ltd. to bring together the two companies and made the assumption that the quality controls we [Baring Brothers] had could get quickly installed there. As it turned out, that appears not to be true."

The most authoritative account comes from San and Kuang (1995: 39) in Singapore. It concludes:

> The organization structure within which Mr. Leeson operated, particularly from the latter half of 1993 through to February 1995 when the Baring Group collapsed, far from embodying some form of integrated control mechanisms with clear lines of accountability, was porous and ill-defined.
>
> In particular:
>
> 1) there was confusion as to the reporting lines and functional responsibilities due to the re-organization of the Baring Group;
>
> 2) the matrix management structure did not work in practice, in relation to Mr. Leeson. Mr. Leeson's local managers took a limited view of their responsibilities. For much of the time Mr. Leeson was trading, it was not clear who his product managers were. Even when this was resolved, his product managers, due to their lack of detailed knowledge of his trading activities, were not able to exercise their functions effectively;
>
> 3) Mr. Leeson was in charge of both the front and back office functions of BFS. Hence, he was in a position (which he exploited) to override internal controls; and
>
> 4) the various departments within the Baring Group, which dealt with matters arising from Mr. Leeson's trading activities were not properly co-ordinated and did not exchange pertinent information which they had in respect of these activities.
>
> If there had been effective checks or organisational controls implemented by qualified personnel, Mr. Leeson would not have been in a position to procure the vast remittances to BFS to finance the trades booked in account 88888, which the senior managers of the Baring Group now claim they know nothing about.

Obviously, the Barings organization was in a period of reorganization, as we define the term, when Leeson undertook his misguided actions. It is noteworthy that the Singapore inspectors ultimately blamed the failure of the bank on the chaotic conditions created by this state, much as our theory suggests. It is also notable that the losses he incurred did not result from a single hidden action or two but, instead, from hundreds of actions sustained over many months.

The reorganization period at Barings might be seen as lasting either ten years (from the establishment of BSL in 1986 to 1995) or six years (from the beginning of the efforts to rein in BSL in 1990). The case for choosing the later date assumes that the two major changes (establishment of BSL and its later reintegration into the bank) were independent events. The case for choosing the earlier date rests on the assumption that the initial decoupled structure at BSL was not fully compatible with the organization's codes—it was just a matter of time until their resolution was attempted. So, the eventual reintegration effort can be seen as a major cascading change. The cultural clashes and conflicts visible for the duration of BSL's life suggest that this is the better interpretation; they also suggest that BSL encountered cultural opposition from the old guard.

The initial autonomy of BSL meant that it was free to adapt to the new business and to the local environment. But it also seems that Barings officials underestimated the potential problems associated with setting up and operating a standalone securities business in Asia. Of course, these officials believed that the information they had was adequate to understand and control what was going on below. Nonetheless, it seems that they failed to foresee two things about control in the new Barings. First, the centuries-old traditional informal control system, based on class homogeneity and background, was ineffective with these new heterogeneous employees. Second, the existing formal information and control systems required transaction-specific knowledge and safeguards to operate effectively. Given their own backgrounds and the bank's history, the leaders lacked this knowledge and failed to appreciate its importance. Officials also seem to have failed to appreciate that the compensation levels and work rules necessary to motivate the entrepreneurial group in Asia on a par with competitors would eventually generate cultural opposition from the more traditional parts of the bank. Tradition at Barings had produced a culture with high asperity that would not readily accept these architectural changes.

Later, opacity in the organization caused leading Barings officials in London and elsewhere to underestimate the costs of the reintegration effort and hence to initiate it. In retrospect, it appears, from the perspective of our theory, that many induced architectural code violations needed to be resolved and that these could not all be undertaken in parallel, as many were interdependent. Yet the bank could not devote all its resources to this effort. It had to continue to operate and to attempt to grow in a competitive environment in the midst of the reorganization.

CONCLUSION

We have attempted here to offer an explanation for why things so often go badly awry when organizations seek to change their architectures. Our theory extends a formal model of cascading organizational change (Hannan, Pólos, and Carroll, 2003a). It examined the implications of opacity, defined as limited foresight about unit interconnections, and asperity, defined as normative restrictiveness to certain architectural features for organizational change. We argued that opacity leads actors to underestimate the lengths of periods of reorganization and the associated costs of change, thereby prompting them to undertake changes with adverse consequences. As with organizational intricacy, both opacity and asperity serve to lengthen the total time that the organization spends reorganizing and increase the associated opportunity costs. The central theorems stated that the likely deleterious effect of a change in architecture on mortality hazards increases with the intricacy of the organizational design, the organization's opacity, and asperity. This theorem has evolutionary implications: intricate, opaque organizations characterized by cultural asperity will tend to get themselves into trouble by undertaking changes that turn out to cost much more than was foreseen—costs that cannot be compensated by the benefits that normally accrue to the new architecture (Hannan, Pólos, and Carroll, 2004).

REFERENCES

Amelio, G.
1998 On the Firing Line: My 500 Days at Apple. New York: HarperBusiness.

Barnett, W. P., and G. R. Carroll
1995 "Modeling internal organizational change." Annual Review of Sociology, 21: 217–236.

Bonacich, P.
1987 "Power and centrality: A family of measures." American Journal of Sociology, 92: 1170–1182.

Carroll, G. R., and M. T. Hannan
2000 The Demography of Corporations and Industries. Princeton, NJ: Princeton University Press.

Crozier, M.
1964 The Bureaucratic Phenomenon. Chicago: University of Chicago Press.

Cyert, R. M., and J. G. March
1963 A Behavioral Theory of the Firm. Upper Saddle River, NJ: Prentice-Hall.

Gamut, L. T. F.
1991 Logic, Language, and Meaning, vol. 2: Intensional Logic and Intensional Grammar. Chicago: University of Chicago Press. (L. T. F. Gamut is the collective pseudonym for J. F. A. K. van Benthem, J. A. G. Groenendijk, D. H. J. de Jong, M. J. B. Stockhof, and H. J. Verkuyl.)

Gross, D., and C. M. Hart
1985 Fundamentals of Queuing Theory, 2d ed. New York: Wiley.

Hannan, M. T., and J. Freeman
1984 "Structural inertia and organizational change." American Sociological Review, 49: 149–164.

Hannan, M. T., L. Pólos, and G. R. Carroll
2003a "Cascading organizational change." Organization Science, 14: 463–482.
2004 "The evolution of inertia." Industrial and Corporate Change, vol. 13: 213–242.

Kahneman, D., and D. Lovallo
1993 "Timid choices and bold forecasts: A cognitive perspective on risk-taking." Management Science, 39: 17–31.

Lager, F.
1994 Ben & Jerry's: The Inside Scoop. New York: Crown Trade.

March, J. G., and H. A. Simon
1958 Organizations. New York: Wiley.

Péli, G., L. Pólos, and M. T. Hannan
2000 "Back to inertia: Theoretical implications of alternative styles of logical formalization." Sociological Theory, 18: 193–213.

Periera, J., and A. Klein
2000 "Returned Xerox chief cites more woes, including bad debts, hotter competition." Wall Street Journal, June 19: A3.

Pólos, L., and M. T. Hannan
2001 "Nonmonotonicity in theory building." In A. Lomi and E. Larsen (eds.), Dynamics of Organizations: Computational Modeling and Organization Theory: 405–438. Cambridge, MA: MIT Press.
2002 "Reasoning with partial knowledge." Sociological Methodology, 32: 133–181.
2004 "A logic for theories in flux: A model-theoretic approach." Logique et Analyse (forthcoming).

Pólos, L., M. T. Hannan, and G. R. Carroll
2002 "Foundations of a theory of social forms." Industrial and Corporate Change, 11: 85–115.

Rawnsley, J. H.
1995 Total Risk: Nick Leeson and the Fall of Barings Bank. New York: HarperBusiness.

San, M. L. C., and N. T. N. Kuang
1995 Barings Futures (Singapore) Pte Ltd: Investigation Pursuant to Section 231 of the Companies Act (Chapter 50). The Report of the Inspectors Appointed by the Minister for Finance. Singapore: Ministry of Finance.

Selznick, P.
1948 "Foundations of the theory of organization." American Sociological Review, 13: 25–35.

Simon, H. A.
1954 "A formal theory of the employment relation." Econometrica, 19: 293–305.

Sterman, J., N. Repenning, and F. Kofman
1997 "Unanticipated side effects of successful quality programs: Exploring a paradox of organizational improvement." Management Science, 43: 503–521.

Stinchcombe, A. L.
1990 Information and Organizations. Berkeley and Los Angeles: University of California Press.

Williamson, O. E.
1975 Markets and Hierarchies. New York: Free Press.

APPENDIX: Theoretical Methodology

In logic, monotonicity means that adding premises to an argument cannot overthrow implications that hold for the simpler premise set. If argumentation is nonmonotonic, then adding more specific premises can eliminate what had been implications in an argument. Pólos and Hannan (2002, 2004) argued that theory building has this nonmonotonic character: better-informed stages of a theory in flux might vitiate implications that were justified in less-well-informed stages.

In formalizing arguments about organizational structure, change, and the hazard of mortality, we used a nonstandard logic designed to represent the nonmonotonic argumentation in theory building (Pólos and Hannan, 2004). This appendix summarizes the key aspects of this language. The distinctive logical component of the language is the definition of four quantifiers: the classical universal quantifier, $\forall$, and three non-classical quantifiers. The first of these, $\mathfrak{N}$, is designed to represent what Pólos and Hannan (2002) called causal stories, rules with possible exceptions. The second, $\mathfrak{P}$, quantifies derived implications, what presumably follows from an argument built partly on causal stories. The third nonmonotonic quantifier, $\mathfrak{A}$, is used in expressing auxiliary assumptions, propositions that are not enduring components of the substantive theory but that allow certain conclusions to be derived from the main postulates.

The language of theory building (Pólos and Hannan, 2004) accommodates three properties that generally characterize theories under construction ("in flux"). Such theories are obviously incomplete. First, they are not what logicians call categorical: some classical first-order sentences might be neither true nor false at a particular stage of the theory. Theory building in this respect involves efforts to eliminate the "unknown" truth-values and to replace them with sentences (formulae) that are either "true" or "false." Second, at least some causal stories are rules with exceptions, rather than universally quantified sentences. Third, a theory in flux might lack some relevant causal stories, and development of the theory might add new rules.

The language of theory building ($\mathfrak{L}_{TB}$) marks the sentences that are responsible for nonmonotonicity with the non-classical quantifiers. The key substantive ingredients of a theory, the specific causal stories, differ from classical first-order (universal) principles in two ways. First, causal stories are more stable than classical principles, which can be overturned by a single counterexample. The language treats causal stories as informationally stable: they are retained even when their first-order consequences are falsified. Instead of being eliminated, the effects of causal stories are controlled (and possibly overridden) by more specific arguments. Second, $\mathfrak{L}_{TB}$ interprets casual stories as what linguists call generic sentences, which express general, but not universal, ideas. The important point is that the truth conditions of a generic sentence cannot be expressed in terms of particular cases; they provide default rules that tell what ought to be expected under normal conditions. The "normally" label for the quantifier $\mathfrak{N}$ signals that the sentence being quantified is a default rule.

Theory building usually involves the commitment to informational stability. As a theory develops, new knowledge adds new causal stories, and the new, more specific rules can override the older, more general rules. This leads to a new stage of the theory. If, however, the new knowledge casts serious doubt on an accepted causal story, then the situation is different. If the new knowledge is dependable, then the old causal story should be discarded. Such a departure from the strategy of maintaining informational stability means that the old theory is replaced by a new one.

$\mathfrak{L}_{TB}$ marks the sentences that can be erased by theory development (the implications of rules with exceptions) with the quantifier $\mathfrak{P}$. Because causal stories are default rules, implications drawn from them on the basis of the best available evidence are defeasible.

Our formalization uses some classical mathematics, especially probability theory and linear algebra. Because the logic of reasoning in these fields of mathematics is classical first-order logic, we developed an interface to combine the classical reasoning and the default reasoning. This interface consists of a special kind of auxiliary assumption. We use the $\mathfrak{A}$ quantifier to mark the auxiliary assumptions.

The interface that matters most in our formalization concerns the role of mathematical expectations. Our probabilistic arguments yield differences in expected value for certain random variables; the causal stories connect the differences of factual values of the same variables to certain other variables. Suppose that a theory stage implies that the expected value of a random variable for one entity is larger than for another: $E\{Y(a)\} > E\{Y(b)\}$. In this case, we argue that the same theory stage should be held to presume that a normal outcome is that $Y(a) > Y(b)$. This consideration yields the following assumption schema:

$$\mathfrak{A}a,b[E\{Y(a)\} > E\{Y(b)\} \rightarrow Y(a) > Y(b)].$$

We call this formula an assumption schema because one can substitute any random variable for Y and any number of equally long sequences for a and b and the resulting formula will be an acceptable auxiliary assumption. It is easy to see that these formulae are not causal stories; they are auxiliary premises that we think make sense in the context. Auxiliary assumptions sit between causal stories and implications (presumptions). $\mathfrak{L}_{TB}$ treats them as assumptions but sets their truth conditions to those of presumptions.

Causal stories refer to regularities in the world. In $\mathfrak{L}_{TB}$, they are true if the claimed regularity is present in the world, and they are false otherwise. This makes their falsification and verification equally difficult, though not impossible. This formal language uses a model-theoretic approach to logic. It builds models for the premises and uses these models to identify the implications of the premises. Premises that provide only partial information, therefore, cannot describe the world completely. Instead of giving details about the actual world, default rules describe several alternative pictures, one of which is the picture of the real world. Logicians call these alternative pictures possible worlds. In contrast, building a model for classical logic needs to refer to only one possible world, the actual one. Logics designed to study arguments that deal with alternative possibilities are called *intensional logics* (Gamut, 1991 provided an accessible overview of developments in intensional logic). The intension of a sentence is a function that tells its truth value in all possible worlds; we denote the intension of the sentence ϕ as $[\![\varphi]\!]$.

A theory stage has two components: a set of possible worlds and a set of regularities (causal stories). The first component captures the factual information: it includes only those possible worlds that satisfy the factual, first-order, premises. $\mathfrak{L}_{TB}$ models regularities as *pairs of open-formula intensions*: the first element in the pair is the antecedent in an implication, and the second is the consequent. Theory augmentations with auxiliary assumptions are represented by theory stages made as similar to the not-augmented theory stage as possible while making the auxiliary assumptions true.

Arguments are modeled by regularity chains. To test whether a theory stage implies a formula of the form

$$\mathfrak{P}[\varphi \rightarrow \psi],$$

one should take the most-specific regularity chains that connect $[\![\varphi]\!]$ to $[\![\psi]\!]$ and demonstrate that at least one of them is more specific than any regularity chain that connects $[\![\varphi]\!]$ to $[\![\neg\psi]\!]$. Here we can restrict ourselves to the first half of the task because the arguments do not supply any regularity chains that connect $[\![\varphi]\!]$ to $[\![\neg\psi]\!]$ in the theorems and lemmas that we prove.