Informal and Formal Organization in New Institutional Economics

ABSTRACT

Exchanges are governed by a set of formal institutions (contracts, incentives, authority) and informal institutions (norms, routines, political processes) which are deeply intertwined. However, for the most part, informal institutions are treated as exogenous forces changing the benefits to use alternative formal structures, and formal institutions are treated as mere functional substitutes of informal elements governing exchanges. As a result, scholars have not sufficiently explored the interactions between formal and informal institutions. We contend that the failure to integrate these concepts into a common theory has led to faulty reasoning and incomplete theories of economic organizations. In this paper, we highlight three potential areas of research exploring the interplay between formal and informal institutions: first, whether formal institutions support (complement) or undermine (substitute for) the contributions of informal institutions; second, how vacillation in formal organizational modes allows managers to efficiently alter the trajectory of informal institutions; and third, how certain informal institutions can lead to hierarchical failure, thereby requiring managers to constrain the boundaries of the firm.
INTRODUCTION

The core argument of new institutional economics (NIE) is that “institutions matter and are susceptible to analysis” (Matthews, 1986, p. 903; Williamson, 1996, p. 3). Institutions provide rules, constraints and incentives that are instrumental to the governance of exchanges. These institutions or governance mechanisms\(^1\) can be either formal or informal in nature (North, 1990). We define formal institutions as rules that are readily observable through written documents or rules that are determined and executed through formal position, such as authority or ownership. Formal institutions, thus, include explicit incentives, contractual terms, and firm boundaries as defined by equity positions. We define informal institutions, in turn, as rules based on implicit understandings, being in most part socially derived and therefore not accessible through written documents or necessarily sanctioned through formal position. Thus, informal institutions include social norms, routines, and political processes.

Despite the recognition by some NIE scholars that informal institutions play a crucial role in defining societal rules (e.g. Denzau & North, 1994; Ensminger, 1997; Greif, 1997), the application of NIE to the study of micro-level issues relevant to business strategy—such as organizational design, firm boundaries, and interorganizational relations—has largely focused on formal institutions. Discussing social norms, Hart (2001, p. 15) contends that “it has been difficult to incorporate norms into the theory of organizations… although there has been some interesting recent work on this topic, this work has not to date greatly changed our views about

\(^1\) Davis and North (1971, pp. 6-7) distinguish between institutions related to the political, legal and social (broadly defined) environment of an economic system, and institutions related to arrangements “between economic units that govern the ways in which these units can cooperate and/or compete.” In this paper, we focus on the latter, which relates to the concept of governance (Williamson, 1991) defined as the “institutional matrix in which the integrity of a transaction is decided” (Williamson, 1996, p. 378, emphasis added). Governance mechanisms, which are comprised of formal and informal institutions, support organizational forms for the production and/or exchange of assets.
the determinants of organizational forms.” Other authors, while acknowledging the role of informal institutions, treat them as exogenous forces that simply change the benefits to using alternative formal structures. For instance, Williamson (1991, p. 359) considers the presence of social networks as a “shift parameter” reducing the incidence of opportunistic behavior and thus favoring non-hierarchical forms of governance. This type of analysis sharply differs from the work of organizational theorists and economic sociologists, who have stressed the central role of informal mechanisms in governing exchanges both internal (Crozier, 1964; Roethlisberger & Dickson, 1939; Trist & Bamforth, 1951) and external to the firm (Granovetter, 1985; Powell, 1990; Uzzi, 1996).

Thus, for the most part, formal institutions have been analyzed and evaluated quite independently of informal institutions. The converse is also true: the study of informal institutions has largely abstracted from the importance of formal institutions, often viewing them as mere functional substitutes. As a result, scholars have not sufficiently explored the interactions between formal and informal institutions. We contend that the failure to integrate these concepts into a common theory has led to faulty reasoning and significant weakness in theories of economic organization. In this essay, we explore how a careful treatment of both informal and formal institutions in the analysis of economic organization provides key insights into the most fundamental predictions of NIE. We submit that a more careful exploration of this relationship is vital to developing a theory of efficient organizational choice.

In particular, we highlight three potential areas of research. First, an improved understanding of the relationship between formal and informal institutions allows us to assess whether formal institutions support (complement) or undermine (substitute for) the contributions of informal institutions. Currently, managers have surprisingly little basis for determining
whether and when formal contracts enhance, damage, or replace trust in interorganizational exchanges or whether and when formal incentives enhance, damage, or replace cooperative behavior within the firm. In this paper, we attempt to provide the beginnings of a theory to untangle this relationship.

Second, a more careful examination of the relationship between formal and informal institutions suggests a need to rethink the fundamental proposition of static alignment pervasive in NIE (Williamson, 1991, p. 277) and other contingency theories in organization theory (Lawrence & Lorsch, 1967). We contend that scholars often mistakenly predict static alignment—matching formal mechanisms to exchange conditions—when the prediction that logically emerges from a more careful assessment of organization theory assumptions is dynamic alignment by vacillation in formal mechanisms. This is because informal institutions affecting the operation of an organizational form, such as political processes and routines, cannot be adjusted directly and thus require changes in formal structures to gradually alter their trajectory. By observing how formal governance causes changes in informal elements, a completely different theory of organizational choice emerges: under some circumstances, even static exchange conditions may demand dynamic responses in terms of formal structures (Nickerson & Zenger, 2001).

Third, a more careful assessment of the relationship between formal and informal governance mechanism is critical to understanding the boundaries of the firm. NIE has successfully built upon Coase’s (1937) seminal work to explain how market failure—namely, transaction costs—shifts the organization of exchanges from markets to hierarchies. However, the theory has not adequately explored the sources of hierarchical failure that pose limits to the
size and vertical scope of firms. We submit that the analysis of informal institutions within firms is critical to understanding how firms determine their boundaries.

This remainder of this paper is organized into five sections. In the first section, we examine the relationship between informal and formal governance and highlight four common assumptions used in the remainder of the essay. The next three sections use these assumptions to explore three ways in which a careful look at the relationship between formal and informal contributes to our development of NIE. Our final section concludes.

THE INTERPLAY BETWEEN FORMAL AND INFORMAL INSTITUTIONS

We argue that the managers’ implicit task is to shape informal and formal institutions influencing the operation of an organizational form in such a way as to increase the functionality that they collectively deliver. By functionality, we mean a variety of dimensions such as capacity to coordinate tasks, to achieve levels of cooperation, or to respond to changing market conditions. In this section, we examine how formal and informal institutions interact and jointly define the functionality of organizational forms.

The role of informal institutions

Research in organization theory has stressed the central role of informal institutions in defining how work is performed and tasks are accomplished within firms (Barnard, 1938; Crozier, 1964; Roethlisberger & Dickson, 1939; Trist & Bamforth, 1951). While formal institutions define the “normative system designed by management” or the “blueprint for behavior” (Scott, 1981, p. 82), informal institutions define the actual behavior of players. Thus, Roethlisberger and Dickson (1939, p. 559) observe that

Many of the actually existing patterns of human interaction have no representation in the formal organization at all, and these are inadequately represented by the formal
organization … Too often it is assumed that the organization of a company corresponds to a blueprint plan or organization chart. Actually it never does (1939, p. 559).

Consistent with this view, many authors remark that the “informal organization,” supported by informal institutions within firms, is not only distinct from formal rules, but that it has a critical role in influencing the operation of firms. Thus, decision making within firms is strongly influenced by political processes (Pfeffer, 1978); patterns of communication are largely a function of informal relationships and shared language (Zenger & Lawrence, 1989); tacit knowledge is rooted in organizational routines (Argote, 1999; Nelson & Winter, 1982); perceived obligations between employer and employee transcend job descriptions and formal contracts (Rousseau & McLean Parks, 1993), just to mention a few examples.

The importance of informal institutions is also recognized in market contexts as well. Granovetter (1985) insists that formal institutions have limited ability to support exchange and thus social networks embodying informal institutions such as norms and trust play a crucial role. Macaulay’s (1963) famous study on the governance of business relationships is consistent with this view. He observes that “businessmen often prefer to rely on a ‘a man’s word’ in a brief letter, a handshake, or ‘common honesty and decency’—even when the transaction involves exposure to serious risks” (1963, p. 58). Thus, informal dealings have the advantage of promoting flexibility and responsiveness to changing conditions, avoiding costly renegotiation of contract clauses (Macneil, 1978). The advantages of such informal contracting mechanisms—commonly referred to as relational governance—are now extensively discussed. Relational governance supports cooperation through norms and reciprocal obligations that transcend initial contract clauses and economize on the costs to use the legal system (Dore, 1983). This discussion leads to:
Assumption 1. Informal institutions strongly influence the functionality of organizational forms.

**Formal institutions as mechanisms of change**

The functional consequences of informal institutions call for managerial action seeking their optimization. For instance, Lincoln (1982, p. 11) observes that “informal networks are indispensable to organizational functioning, and managers must learn to manipulate them for organizational ends.” However, the fact that informal institutions in general are difficult to manipulate engenders major managerial challenges. Fortunately, changes in formal institutions, which can be directly manipulated, appear to strongly influence changes in informal institutions within firms. Internal routines, norms, and networks of influence develop over time in response to an organization’s formal structure (Shrader, Lincoln, & Hoffman, 1989; Stevenson, 1990; Tichy, 1980). Research shows, for instance, that the operation of informal networks is influenced by the positions of individuals in the formal hierarchy (Brass, 1984; Krackhardt, 1990).

Formal institutions also appear to influence the trajectory of informal elements in interorganizational relationships. Long-term contracts and joint equity stakes (ownership) can create strong commitments between parties and thereby promote the emergence of mutual trust (Doz, 1996; Parkhe, 1993). But changes in formal institutions can also be used to disrupt dysfunctional informal ties between representatives of transacting firms. For instance, Humphrey and Ashforth (2000, p. 719) document that U.S. automakers expressed “concerns that interpersonal relationships could lead buyers to award contracts to suppliers who had higher unit costs, lower quality and slower delivery times.” The adoption of competitive bidding—such as in the case of “business-to-business” exchanges through the Internet—is seen as an opportunity
to circumvent reciprocal deals between sales and buying representatives which may not be in the best interests of one employer or the other.

Therefore, since “the formal largely orders the direction the informal takes” (Dalton, 1959, p. 237), formal institutions constitute a tool available to managers through which informal institutions can be shaped. This supports:

Assumption 2. Formal institutions influence the trajectory of informal institutions.

The nature of formal and informal changes

Several organizational perspectives share the assumption that formal institutions involve discrete modes comprised of “bundles” of mutually consistent, complementary features. In discussing TCE, Williamson (1991, p. 271) stresses that “… each viable form of governance… is defined by a syndrome of attributes that bear a supporting relation to one another.” Organizational economists Milgrom and Roberts (1991, p. 84) also submit that organizations involve activities that are “mutually complementary and so tend to be adopted together with each making the others more attractive.” For instance, centralization is characterized by structural interdependence between units, lower-powered incentives, and centralized decision making; decentralization is characterized by structural autonomy, higher-powered incentives, and local decision making. Each element reinforces the other: for instance, the use of higher-powered incentives is expected to discipline autonomous units to act efficiently, while autonomy supports those incentives since performance is assessed on an individual basis. On the other hand, the use of one element in isolation or in conjunction with another incompatible element will yield sub-optimal results. Thus, the adoption of higher-powered incentives jointly with centralized decision making is likely to trigger dysfunctional attempts by local managers to influence the central manager’s decisions or alter performance standards.
The assumption that organizational forms are discrete is also pervasive in organization theory. The configuration literature considers that organizational structures are composed of clusters of consistent traits (Meyer, Tsui, & Hinings, 1993; Mintzberg, 1979). Miller and Friesen’s (1980, p. 593) empirical study evidences that changes in organizational variables “tend to occur together, or … follow one another after a very brief interval, in order to maintain an appropriate balance or ‘configuration’.” Punctuated equilibrium models describing organizational change share the same perspective. Gersick (1991) use the term “deep structure” to describe systems with distinct and interdependent parts, which change in an abrupt, comprehensive fashion rather than gradually. Likewise, Tushman and Romanelli (1985) consider that change is followed by periods of convergence that align the diverse activities within a firm into a consistent portfolio. Change (or reorientation) causes consistent changes of these activities toward a new alignment or deep structure, which clearly suggests discrete choices.

In contrast with formal organizational structures, which correspond to menus of discrete choices, informal elements are continuously arrayed. Social attachments, for instance, differ in degree rather than in kind. Thus, Granovetter (1973, p. 1361) defines tie strength as a combination of the “amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie.” These elements clearly have continuous flavor. Talking about individual commitment to an organization, Salancik (1977, p. 4) points out that “there are degrees of commitment [which derive] from the extent to which a person’s behaviors are binding.” Krakhardt (1990) uses the continuous measure of individual centrality in a social network to indicate the degree of an individual’s power within an organization, derived from his or her ability to control information flows.
Thus, while changes in formal institutions are expected to involve discrete, abrupt movements of consistent variables, changes in informal institutions have a more continuous and gradual character. This leads to:

Assumption 3. Formal institutions are discretely arrayed, while informal institutions operate comparatively on a continuum.

The pace of formal and informal changes

Although changes in formal institutions trigger changes in informal institutions, the latter do not respond instantaneously. The concept of inertia, pervasive in organization theory, implies that webs of interdependent relationships, political coalitions, patterns of communication, established routines impede organizational change (Hannan & Freeman, 1984; March & Simon, 1958; Nelson & Winter, 1982; Tushman & Romanelli, 1985). The existence of inertia causes the functionality of an organizational form to change slowly, thereby creating a lag between the implementation of a new formal structure and the change in the overall functionality, which derives in large part from informal elements.

The concept of inertia is also pervasive in NIE. Building upon Arthur’s (1989) and David’s (1985) ideas, North (1990) discusses how institutions exhibit *path dependence* in that the trajectory of an economic system is largely a function of its past position. North attributes a great deal of such path dependence to informal institutions derived from “available mental constructs—ideas, theories, and ideologies” (1990, p. 96), which create resistance to change. In the same vein, Greif (1997, p. 89) argues that “past behavior, cultural beliefs, social structures, and organizations impact the development of values and social enforcement mechanisms that inhibit flexibility in departing from past patterns of behavior.” A common theme in the NIE

---

2 See also Denzau and North (1994).
literature discussing economic change is that although it is easy to implement changes in formal institutions (laws, decision rights, etc.), existing informal institutions are difficult to disrupt, responding gradual and slowly to formal changes. Thus:

**Assumption 4.** Formal and informal institutions differ in the pace with which they change.

- Informal institutions possess inertia that slows the pace of change.

In the sections that follow, we discuss how the assumptions presented above imply three general propositions based on the relationship between informal and formal institutions:

**Proposition 1.** Formal and informal institutions are interdependent governance mechanisms in that the use of one mechanism can either promote (complement) or undermine (substitute for) the use of the other.

**Proposition 2.** Even in static environments, achieving the optimal functionality of an organizational form may require dynamic changes in formal institutions. Thus, under some circumstances, a pattern of vacillation in formal institutions supporting distinct organizational forms (market vs. hierarchy, centralized control vs. decentralized control, etc.) is warranted (Nickerson & Zenger, 2001).

**Proposition 3.** Firm boundaries are determined in large part by the need to adjust informal institutions within hierarchies. In particular, managers must sever the boundary of the firm to suspend dysfunctional informal processes.

Each general proposition is explained in turn. Taken together, these propositions exemplify how a more careful examination of the relationship between formal and informal institutions may provide important insight to our understanding of organizational choice.
INFORMAL AND FORMAL INSTITUTIONS: COMPLEMENTS OR SUBSTITUTES?

If informal institutions influence the functionality of an organizational form (Assumption 1) and their trajectory is determined in part by formal institutions (Assumption 2), then a key question centers on the nature of relationship between these two mechanisms. At the most basic level, one must ask whether the use of one type of institution increases or decreases the functionality of the other—i.e., whether informal and informal institutions function as complements or substitutes. As it turns out, the literature focusing on this issue diverges along several paths.

Formal and informal institutions as complements

*Substitution* arguments cluster around two basic claims. A mild perspective, which we refer to as *weak* substitution, argues that formal constraints are unnecessary because informal relationships based on trust and social norms can support cooperation without the costs and complexity incurred with formal agreements (Ellickson, 1991; Gulati, 1995; Powell, 1990; Ring & Van de Ven, 1994; Uzzi, 1996). Granovetter (1985, p. 489) contends that formal institutions “do not produce trust but instead are a functional substitute for it.” According to this view, social norms support the emergence of trust because buyers can hope to get some cooperation even when formal instruments are absent. One of the most discussed social norms is *reciprocity*, meaning that individuals tend to cooperatively respond to generous offers even if this is against their own self-interest (Berg, Dickhaut, & McCabe, 1995; Dore, 1983; Rabin, 1993). Thus, a buyer can reciprocate a high-quality service by offering an above average price to a seller even if they are not expected to meet one another in the future; anticipating this act of reciprocity, the seller will be motivated to supply a high-quality service in the first place. As Hoffman, McCabe and Smith (1998, p. 338) put it, reciprocity functions as an enforcement mechanism because it
“leads naturally to property rights” (emphasis in the original). Thus, weak substitution implies that the presence of informal institutions such as norms and trust reduce the benefits to use formal institutions; the latter become simply unnecessary.

A starker substitution perspective, which we refer to as strong substitution, argues that formal institutions are not only unnecessary but also damaging to the formation and operation of informal elements. Macaulay (1963, p. 64) contends that “not only are contract and contract law not needed in many situations, their use may have, or may be thought to have, undesirable consequences… Detailed negotiated contracts can get in the way of creating good exchange relationships between business units.” He argues that some firms discourage the use of an elaborate contract because it “indicates a lack of trust and blunts the demands of friendship, turning a cooperative venture into an antagonistic horsetrade” (Macaulay, 1963, p. 64).

Similarly, Sitkin and Roth (1993, p. 376) posit that “legalistic remedies can erode the interpersonal foundations of a relationship they are intended to bolster because they replace reliance on an individual’s ‘good will’ with objective, formal requirements.” Ghoshal and Moran (1996) also stress that the use of rational, formal control has a pernicious effect on cooperation. They contend that for those parties being controlled…

… the use of rational control signals that they are neither trusted nor trustworthy to behave appropriately without such controls. … For the controller, negative feelings arise from what Strickland (1958) described as ‘the dilemma of the supervisor’ viz., the situation when the use of surveillance, monitoring, and authority led to management’s distrust of employees and perceptions of an increased need for more surveillance and control… (1996, p. 24)

Social psychologists have provided an explanation for this effect: explicit incentives or punishments may reduce partner’s intrinsic motivation to perform certain tasks (Deci & Ryan, 1971). Interestingly, Williamson (1996, p. 271) makes a similar argument, but restricts its application to purely social, non-economic relationships.
This effect has received the name *motivation crowding out* in economics (Frey, 1997). According to proponents of motivation crowding out theory, reciprocity is a particular form of intrinsic motivation originated from social norms that is violated when formal incentives or punishments are present (Gächter & Falk, 2000). In other words, those mechanisms can signal that trust is absent and no reciprocity is expected, thereby framing the relationship in a strictly economic, rather than social, orientation (Lubell & Scholz, 2001; Tenbrunsel & Messick, 1999). One possible consequence is that the outcomes achievable through incentives and controls can be less efficient than those that could naturally flow from an individual’s voluntary willingness to cooperate, manifested through social norms and trust.

**Formal and informal institutions as complements**

An alternative argument that has received comparatively less attention is that formal institutions complement informal mechanisms. In settings where hazards are severe, the combination of formal and informal safeguards may deliver greater functionality than either institutional type in isolation. As North (1990, pp. 46-47) puts it, “formal rules can complement and increase the effectiveness of informal constraints. They may lower information, monitoring, and enforcement costs and hence make informal constraints possible solutions to more complex exchange.” Poppo and Zenger (forthcoming) provide supportive empirical evidence. They find that customized contracts appear to support relational governance, characterized by alignment of goals, trust and collaborative orientation.

Three distinct arguments support the complementarity proposition. First, formal contracts may both extend the expected duration of a relationship and restrict the gains from one-time deviations from cooperative behavior in an exchange relationship (Baker, Gibbons, & Murphy, 1997). Contracts not only have this source of advantage because of their formal
specification of a long-term commitment to exchange, but also can limit the domain and lessen
the gain of potential opportunistic behavior through clearly articulated clauses that specify
punishments. This reduction in short run gains heightens comparatively the gains from
cooperating in the exchange relationship. By contrast, failing to contractually specify elements
of the exchange that are easily specified merely heightens incentives for short-run cheating and
lowers expectations of cooperation (Baker, Gibbons, & Murphy, 1994; Klein, 1996; Milgrom,
North, & Weingast, 1990). Note that complementarity arguments assume that formal institutions
are to some extent incomplete, since otherwise any outcome could be legally enforced without
the need of informal institutions. Due to the costs to write clauses, limits of enforceability by
courts and individuals’ cognitive limitations (bounded rationality), it is not possible for parties to
pre-specify all future contingencies in a comprehensive contract.4

Lazzarini, Miller and Zenger’s (2001) experiment provides support for the
complementarity argument outlined above. Their experiment involves buyer-seller exchanges
with moral hazard on the part of sellers. To operationalize contract incompleteness, the authors
consider that the good being transacted has two distinct dimensions. One dimension is easy to
specify and therefore is contractible in advance: buyers can structure contingent payments based
on the supplied level of that dimension. The other dimension is difficult to measure and enforce
by third parties (such as courts), and thus is non-contractible: no contingent payment can be
applied. Lazzarini, Miller and Zenger (2001) find that contractual incentives (contingent
payments) applied to the contractible exchange dimension facilitate the enforcement of the non-
contractible dimension, precisely because they limit the gains that sellers could attain from short-
term defection.

---

4 For a recent assessment of the incomplete contracting literature, see Tirole (1999).
The second argument providing support for the complementarity view is that formal institutions can set the stage for the development of trust within a long-term interaction. This is a direct implication of Assumption 2: formal mechanisms can influence the trajectory of informal elements. Cooperative behavior in the present—as a result of supporting formal mechanisms—reinforces an expectation of cooperation in the future. Supportive of this logic, empirical work suggests that past success in contracting with a particular exchange partner yields greater success in the present (Anderson & Weitz, 1989; Larson, 1992). Formal contracts help ensure that the early, more vulnerable stages of exchange are successful. Durkheim (1933) appears to have this idea in mind when he writes that

… in order for [parties] to co-operate harmoniously, it is not enough that they enter into a relationship, nor even that they feel the state of mutual dependence in which they find themselves. It is still necessary that the conditions of this co-operation be fixed for the duration of their relations. The rights and duties of each must be defined, not only in view of the situation such as it presents itself at the moment when the contract is made, but with foresight for the circumstances which may arise to modify it. Otherwise, at very instant, there would be conflicts and endless difficulties (1933, pp. 212-213).

Thus, formal institutions may also be designed to create procedures to adapt to changing conditions. Unexpected disturbances may place considerable strain on an exchange relationship (Williamson, 1991, pp. 271-273). Formal contracts that shift from merely specifying deliverable outcomes to providing frameworks for bilateral adjustments may facilitate the evolution of highly cooperative exchange relations. Crocker and Masten (1991, p. 95) suggest that “it seems more appropriate to view contracts as means of establishing procedures for adapting exchange and resolving disputes rather than purely as incentive mechanisms.” In addition, the process of contracting may itself promote expectations of cooperation consistent with relational governance. The activity of creating complex contracts requires parties to mutually determine and commit to processes for dealing with unexpected changes, penalties for non-compliance, and
other joint expectations of trade. Hence, the process of developing complex contracts in response to exchange hazards positively affects future exchange performance through the development of social relations resulting from the very act of bilaterally negotiating contract terms.

The third argument supporting the complementarity view is that informal elements may also promote the refinement (and hence increased complexity) of formal institutions. As discussed before, informal institutions increase the performance of formal institutions because explicit arrangements are inherently incomplete. Thus, not only do formal institutions promote the stability of informal institutions, but informal institutions also play a role in filling contractual gaps over time. As a close relationship is developed and sustained, lessons from the prior period are reflected in revisions of the contract. Exchange experience, patterns of information sharing, evolving performance measurement and monitoring may all enable greater specificity (and complexity) in contractual provisions and exchange conditions. As a consequence, relational exchanges may gradually develop more complex formal contracts, as mutually agreed upon processes become formalized.

An integrated assessment

We note that the substitution and complementarity effects described above are not mutually exclusive; the interaction between formal and informal institutions is too complex to accommodate a unique pattern. For instance, even if explicit incentives or control mechanisms reduce individuals’ intrinsic motivation to provide extra effort, they can at the same time discourage short-term defection. Thus, the effects described above can be best viewed as partial effects; the net outcome is dependent on particular exchange conditions. Although research on this topic is on its early stages, we tentatively outline some specific propositions. We submit that
formal and informal institutions act more as complements than substitutes when (1) individuals are not likely to or not committed to transact repeatedly and (2) the procedures involved in the operation of formal institutions are perceived to be “fair”. When these conditions do not hold, the use of formal institutions may be unnecessary and even detrimental to the operation of informal institutions.

The reason for our first claim is that non-repeated interactions provide neither a “shadow of the future” increasing individuals’ perceived benefits from cooperation (Axelrod, 1984) nor a “shadow of the past” promoting the gradual development of relational norms and trust (Macneil, 1978; Ring & Van de Ven, 1994). Hence, the benefit of formal institutions becomes relatively more important in new or non-recurring relationships, since informal enforcement will tend to be weak or absent (Lazzarini et al., 2001). One could argue that, instead of relying on formal structures to support non-repeated exchanges, parties might be better off encouraging repeated interaction as a way to create norms and trust (e.g. Dyer & Singh, 1998; Kollock, 1994; Krackhardt, 1992). However, parties must credibly commit to a repeated interaction since at any moment they can switch to alternative partners. Formal arrangements are thus a way to lock parties into relationships with sufficient duration (Baker et al., 1997). In addition, repeated interaction between the same agents limits the opportunities and information that they can attain with external relations (Blau, 1964). Non-recurring exchanges or “weak ties” are a fundamental way to transfer new information and knowledge between specialized agents (Granovetter, 1973). By contrast, “overembedded” systems are likely to involve low knowledge diversity and hence less propensity to innovate (Greif, 1997; Uzzi, 1996). Formal institutions are thus crucial for economic growth marked by specialization and fewer recurring exchanges (North, 1990; Zucker, 1986).
Our second claim derives from the growing recognition that individual attitudes are dependent on procedural issues, which alter their perceptions about the fairness of processes employed by parties to achieve certain outcomes (Bies & Shapiro, 1988). Consider for instance the effect of contract “framing”: empirical studies provide evidence that people tend to prefer equivalent contracts that specify rewards or “bonuses” rather than punishments or “damages” (Luft, 1994; McLean Parks & Coelho-Kamath, 1999). It is possible that equivalent contracts stipulating bonuses instead of punishments discourage defection without crowding out implicit norms. In addition, the procedural perceptions of formal institutions are dependent on the extent to which parties expect the use of these institutions for certain types of exchanges, i.e., whether their use is “taken for granted” or not. Thus, Lewicki, McAllister, and Bies (1998, p. 454) argue that “quality control people do their work because it’s their job—not necessarily because they personally distrust others.” Also, two firms engaging in an alliance may employ a formal contract not because they do not trust one another in particular, but because it is a standard procedure in their industry. Thus, whether a formal institution complements or substitutes for informal institutions depends in part on how fair that formal institution is perceived to be.

In conclusion, formal and informal institutions are not mere alternative ways to govern exchanges. In most cases they are employed simultaneously and interact in complex ways. The complete assessment of complementarity and substitution effects, and in which conditions one effect supplants the other, is an important research agenda both within and outside NIE.

**THE DYNAMIC ALIGNMENT OF ORGANIZATIONAL FORMS**

Theories within NIE argue that that the choice of organizational forms involves matching formal structures to strategies, exchange conditions, and environments in some discriminating
way (Chandler, 1962; Williamson, 1975). NIE is fundamentally a theory of static alignment and, as such, shares many of the elements of “contingent-fit” models. Namely, selection and external pressures (e.g., from owners and capital markets) will either prompt managers to choose organizational forms that are aligned to particular exchange conditions, or wash out misaligned forms. Thus, managers choose to govern exchanges through markets when outputs are easily measured (Barzel, 1982; Holmstrom & Milgrom, 1994; North, 1981), and the production of those outputs involves low levels of specialized assets (Klein, Crawford, & Alchian, 1978; Williamson, 1985). By contrast, managers choose hierarchy when outputs are difficult to measure and specialized assets are substantial. Intermediate conditions may favor “hybrid” organizational forms such as long-term contracts and interfirm alliances (Williamson, 1991).

Similar discriminating alignment logic governs the choice of governance forms within the firm. Thus, managers may choose decentralized forms with their incumbent high-powered incentives when innovation is desired, but adopt centralization with its ready access to authority when coordination is desired (Williamson, 1985).

Static alignment implies that if the diverse contingencies that affect firms remain stable, then organizational forms should also remain unchanged. Changes in formal mechanisms are precipitated by changes in environment or exchange conditions, which are influenced in part by strategy decisions. The patterns of change in formal governance that we commonly observe, however, are often difficult to reconcile with this theory of static alignment. Many changes in formal governance appear to occur with little change in environment or exchange conditions. Indeed, we seem to often observe firms engaged in vacillating patterns of choice in formal organization (Carnall, 1990, p. 20; Cummings, 1995, p. 112; Eccles & Nohria, 1992, p. 127; Mintzberg, 1979, p. 294). Thus, firms centralize, then decentralize, then centralize, etc. Firms
outsource an activity, then internalize it, only to outsource it, again. In the choice of compensation, many firms seem to vacillate between aggressive incentive pay for sales personnel and flat salaried pay.

Examples of vacillation abound (Nickerson & Zenger, 2001). Six times within 16 years, Hewlett-Packard made fundamental shifts between centralizing and decentralizing core activities within the corporation, without any change in exchange or environmental conditions. Consider also KPMG Peat Marwick. Prior to 1992, the consulting and accounting services firm was structured geographically around local managing partners, which helped to leverage close ties with regional clients into a broad range of services. In 1992, KPMG shifted to a functional structure where associates reported to nationwide practice managers rather than local managers. This structure promoted knowledge sharing and resource allocation within particular consulting and accounting practices. In 1994, KPMG shifted again its organizational configuration to an industry-focused structure based on categories such as healthcare, government services, retail, and manufacturing. This new structure promoted the development of in depth knowledge about client industries. In 1996, KPMG returned to its former geographic structure centered on local managing partners. However, by 1999 they shifted back to an industry-focused industry, although now globally centralized. Within 7 years, KPMG had vacillated among three (discrete) formal structures—geographic, functional, and industry-focused—with 5 events of structural change. How do we explain this pattern, since it is very unlikely that within this short period KPMG had faced so many changes in exchange and environmental conditions?

Arguably, such vacillation in formal organizational modes can result from managers’ fickle behavior, thereby being simply a manifestation of noise in the decision making process. However, as Nickerson and Zenger (2001) maintain, a richer understanding of the relationship
between formal and informal institutions leads to the conclusion that such commonly observed vacillation can indeed be efficient. The explanation for functional vacillation follows logically from the assumptions described in section 2.

As indicated by Assumption 1, the functionality of an organizational form is determined as much by informal mechanisms, as by formal mechanisms. In addition, informal institutions vary systematically in response to changes in formal institutions (Assumption 2). But although managers can influence changes in informal institutions, they only have discrete formal choices or “levers”—e.g., centralization vs. decentralization, make vs. buy, etc.—to promote such changes (Assumption 3). If the desired level of functionality—which is largely dependent on informal institutions—lies in between the functionality delivered by these formal levers and change is not too costly, then managers will have an incentive to modulate between two or more discrete formal choices to achieve temporarily the desired intermediate level. Given that informal institutions display inertia (Assumption 4), each switch between formal choices triggers a gradual change in the trajectory of informal elements. Thus, by vacillating between distinct formal choices, managers can influence the trajectory of informal institutions towards a desired position that is unavailable if the organization remains fixed with a particular formal structure.

The choice between centralization and decentralization is a useful illustration of the virtues of vacillation. Centralization and decentralization are discrete organizational modes characterized by distinct sets of formal institutions. While centralization involves structural interdependence between units, lower-powered incentives, and centralized authority, decentralization involves structural autonomy, higher-powered incentives, and local authority. These organizational modes also exhibit distinct and conflicting patterns of functionality. Centralization facilitates coordination, but at the cost of low-powered incentives and reduced
innovation potential. Decentralization yields high-powered incentives and increased innovation, but at the cost of coordination. Managers would like to maintain a level of functionality such that improved coordination and higher-powered incentives/innovation coexist; by choosing permanently either centralization or decentralization, managers will necessarily sacrifice one of these dimensions of functionality. They can, however, dynamically modulate between these two formal structures to achieve temporarily a level of functionality that lies in between centralization and decentralization.

Thus, a decision to change from centralization to decentralization triggers changes in informal elements that alter the functionality of the organizational form. As managers initiate decentralization and the firm begins to reap benefits from higher-powered incentives and innovation, it still enjoys, albeit temporarily, the dense communication channels and social attachments supported by the informal institutions (e.g., routines, norms, etc.) that accompanied the formerly centralized structure. However, these patterns will tend to diminish as time elapses, since the formal change to decentralization will sever social attachments and communication flows within the firm, which in turn will dampen coordination. As a result, the overall organizational performance will migrate towards the steady-state functionality delivered by decentralization. When this occurs, managers can do the reverse: they can change the formal structure from decentralization to centralization to restore coordination while keeping some innovation and incentives reminiscent from the decentralized structure. However, after some time, sticky routines and excessive politicking promoted by centralization will cause again a reduction in functionality. Managers will then need to initiate a new cycle by decentralizing the organization in order to alter these informal elements.
Nickerson and Zenger (2001) show that this theory provides a counterintuitive result that \textit{inertia can be performance enhancing}. The reason is that inertia reduces the frequency with which managers must change the formal structure. When inertia is high, informal institutions change slowly in response to changes in formal institutions. Since informal elements are critical determinants of organizational functionality, the latter will remain close to the optimal level—which will combine elements of both discrete forms—for a long period of time before a new switch is necessary. Were changes in informal institutions instantaneous (i.e., no inertia), it would not be possible to keep informal institutions at the desired, intermediate level for a time period sufficient to warrant vacillation: they would quickly converge to the steady-state position of the chosen formal structure.

Thus, the manager’s task is not simply to observe changes in environment or exchange conditions, but rather to monitor the trajectory of informal institutions and manipulate them indirectly through changes in formal structures. Vacillation between organizational forms comprised of discrete formal institutions is efficient when the costs of change are moderate and the desired functionality lies intermediate to that delivered by either formal form in steady state, since informal institutions will lag formal changes and therefore will stay temporarily at that intermediate position. Hence, taking informal institutions seriously has in this case led us to a conclusion that contradicts a fundamental proposition from NIE (and TCE in particular): rather than statically align institutions to exchange or environmental conditions, under certain circumstances managers must pursue a dynamic alignment by vacillating between discrete formal institutions. In this sense, changes in formal structures can occur even when exchange or environment contingencies remain unchanged.
INFORMAL INSTITUTIONS AND FIRM BOUNDARIES

The prototypical institutional choice in NIE is the choice between market governance and hierarchical governance. This is fundamentally viewed as a choice between two formal institutions. The standard story—derived from Coase’s (1937) insight—is that the institution of hierarchy is chosen when markets fail. Thus, we have well developed explanations for why markets fail and therefore why managers choose to replace markets with internal organization. We have also have a large body of confirmatory empirical evidence: managers appear to choose hierarchy when exchange conditions present hazards in using markets and choose markets when they do not (e.g. Poppo & Zenger, 1998; Shelanski & Klein, 1995). Managers choose hierarchy because hierarchy possesses governance features to which markets have limited access. Namely, hierarchy’s low-powered incentives discourages expropriation of rents in the presence of specific assets (Williamson, 1985) and avoids dysfunctional responses to incentives when performance attributes are difficult to measure (Holmstrom & Milgrom, 1994).

However, as numerous scholars have noted, why managers choose markets is not as well understood. Parallel logic suggests that markets are chosen when hierarchies as institutions fail. Interestingly, Coase (1937) in his seminal paper asks

… why, if by organizing one can eliminate certain costs and in fact reduce the cost of production, are there any market transactions at all? Why is not all production carried on by one big firm? (1937; reprinted in 1991, p. 23)

Coase tentatively answers this question by invoking, among other things, the nebulous concept of “diminishing returns to management,” prevalent in early industrial organization writings, which asserts that managers have limited ability to coordinate large flows of resources. This explanation is not satisfactory because, for instance, one could solve the problem of diminishing returns by splitting the firm into smaller independent units, and then using internal
markets to allocate resources. At a more fundamental level, we must ask why hierarchies cannot selectively use the high-powered incentives of markets and thereby do all that markets can and more (Williamson, 1985, p. 133). What are the limits to using market-like instruments to reduce the deficiencies of hierarchies?

Within this perspective, Williamson (1985) provides a more convincing explanation of the limits of firms: it is not possible to “selectively intervene” inside hierarchies by infusing market-like incentives without incurring additional costs or creating undesirable side effects. He notes, for instance, that performance indicators can be manipulated, and incentives (such as piece rates) can lead to the overutilization of a firm’s assets by employees. Williamson (1985, p. 142) also briefly mentions the role of fairness considerations in dictating the allocation of gains and losses inside firms, which are rooted in informal institutions. This, however, is only the tip of the iceberg. We submit that a thorough understanding of informal institutions within organizations is critical to understanding the limits of the firm and to developing a theory of firm boundaries.

The basic argument is as follows. A key reason why hierarchies reduce market failure is that they trigger the formation of informal institutions—norms, routines, organizational culture, etc.—which affect organizational functionality (Assumption 1) by facilitating communication, coordination and cooperation (Barnard, 1938; Kogut & Zander, 1996). However, such informal institutions also bring side effects. For instance, social attachments cause biased decision making; firm-specific routines constrain the ability of those within the firm to externally communicate and acquire external knowledge. Thus, the problem that managers of hierarchy face is that the informal institutions which hierarchy triggers cannot be selectively shut down with any great success. Consequently, to suspend these informal processes, managers must shift
or constrain the boundaries of the firm. Since these informal processes are influenced by formal decisions (Assumption 2), we contend that firms can adjust their formal boundaries to alter the dynamics of informal institutions. Thus, firms suspend hierarchy as an institution (i.e. sever the organizational boundary) to avoid the informal processes that run rampant within their boundaries. This shows that, in taking seriously the concept of informal institutions, we are able to develop a theory of hierarchical failure—a theory that allows us to explain why firms constrain their boundaries. While informal features of hierarchies reduce market failure, they create costs that need to be factored in boundary choices.

We identify here four critical informal processes, largely based on informal institutions, creating hierarchical failure: (1) influence activities, (2) social attachments, (3) social comparison processes, and (4) development of firm-specific routines. We discuss each in turn.

**Influence activities**

Milgrom and Roberts (1990) argue that influence activities—attempts to influence the allocation of rents within firms in order to preferentially reward particular individuals or coalitions—are the primary costs of internal organization. Hierarchies not only create an environment where individuals can engage in lobbying to distort the allocation of resources, since exchanges are not disciplined by market forces, but also magnify the extent to which such activities are feasible. This is because the increased communication channels within hierarchies, although instrumental in facilitating coordination (Eccles & White, 1988; Pfeffer, 1978), represent ways in which individuals can reach and influence decision-makers. In this sense, politicking becomes by itself a fundamental informal institution governing and affecting the functionality of hierarchies.
Poppo (1995) provides evidence on how influence activities have implications for boundary decisions. She finds that although hierarchy improves coordination by facilitating the exchange of information between internal units, it engenders difficulty in negotiating internal (transfer) price adjustments due to costly bargaining among divisional managers. Consequently, since the coordination-based benefits of hierarchy necessarily unleash influence activities, firms must constrain their boundaries to interrupt communication channels that facilitate such dysfunctional political behavior.

**Social attachments**

Similar to influence activities, social attachments can distort the allocation of resources within firms, which in turn affects the costs of internal organization. Thus, poor decision and resource misallocation may occur even in the absence of lobbying efforts or other influence activities. Such social attachments are largely governed by a host of informal institutions. Thus, decision-makers may overfund projects or divisions with rather limited promise in an act of reciprocity to friends/managers, or underfund projects and divisions with more substantial promise but involving managers with whom they lack such social attachments. Evidence also suggests that reciprocity norms embedded in friendship ties can reduce individuals’ willingness to negotiate freely and pursue better opportunities (Halpern, 1994). In addition, excessive socialization can induce a bias towards shared perspectives, values, and culture, thereby undermining the firm’s ability to find innovative solutions (Gruenfeld, Mannix, Williams, & Neale, 1996; Katz, 1982). Thus, to the extent that social attachments within firms lead to biased decision making, managers may wish to limit their boundaries in order to reduce the reach and consequences of these distortions. Resource allocation decisions governed through the market are likely to be less contaminated by such distortions, precisely because they will be less affected.
Social comparison processes

The social norm of equity in the allocation of rewards creates difficulty in the design of compensation schemes within firms (Adams, 1965; Deutsch, 1985). Employees directly, or indirectly through managers, compare their rewards to all other employees within the boundary of the firm. When they perceive rewards to be inequitably distributed—i.e., when they consider that their compensation to effort ratio is lower than other colleagues—they reduce effort, seek to alter the distribution, or simply depart the firm (Adams, 1965). All such outcomes are costly to the firm. The challenge that a manager faces in rewarding employees is that employees possess highly inflated perceptions of their own performance, exacerbated by the fact that the manager does not possess a fully accurate measure of performance. Hence, efforts to aggressively reward performance in the absence of accurate performance measures trigger social comparison processes that impose costs upon the firm. In response to these social comparison processes triggered by equity norms, managers simply adopt low-powered performance incentives. However, this brings two side effects (Zenger, 1994). First, low-powered incentives are likely to reduce employees’ effort compared to the situation involving high-powered incentives. Second, other firms offering contracts with a closer match between pay and performance will lure more skilled people, since the latter will be able to reap higher rents from their superior talent. Thus, firms offering low-powered incentives are likely to face not only turnover costs, but also the departure of skilled people to other firms (Zenger, 1992).

Constraining the boundaries of the firm and more specifically reducing the size of the firm constrains the scope of social comparison processes. The costs of social comparison that are associated with a high-powered incentive scheme are strongly dependent on the size of the firm. In large firms there is a much larger group of individuals who will respond negatively
when a colleague within the firm is granted a significant performance-based increase in pay. In small firms, the number of comparisons triggered by a single adjustment in pay is substantially less. Further, in small firms, information about individual performance levels is more easily disseminated. Consequently, a reduction in firm size is likely to enhance consensus about individuals’ perceived relative performance and hence reduce the costs created by social comparisons. It follows that small firms will tend to offer higher-powered incentives than large firms, since they will face lower inequity perceptions resulting from such incentive policies. For this reason, managers may sever the boundaries of the firms to shut down, at least in part, social comparison processes created by equity norms within hierarchies.

**Firm-specific routines**

Firm-specific routines develop as a by-product of repeated interaction within firms, and represent informal institutions in the form of tacit codes of interaction (Nelson & Winter, 1982). While some consider such common vocabulary and procedures as enhancing communication (Allen & Cohen, 1969; Lawrence & Lorsch, 1967), it does not come without cost. Tushman and Katz (1980) point out that

> The evolution of local languages and coding schemes helps the unit deal with its local informal processing requirements; yet, it also hinders the unit’s acquisition and interpretation of information from external areas. External communication is vital, however, both in terms of feedback and for evaluating and acting on the unit’s environment (1980, p. 1072).

Thus, idiosyncratic routines can lock partners into one particular field of knowledge and lock them out of external opportunities and sources of information (Cohen & Levinthal, 1990; Leonard-Barton, 1995; Poppo & Zenger, 1998). Employees sharing common vocabulary and interpretations are not only likely to have limited capacity to interpret external sources of
information, but also to recognize the importance of those sources (Levinthal & March, 1993). As Katz (1982) explains,

One of the main principles of human communication… is the strong tendency for individuals to communicate with others who are most like themselves or who are most likely to agree with them. Over time, project members learn to interact selectively or avoid messages and information that might conflict with their established practices and dispositions, thereby reducing their overall levels of outside contact (1982, p. 85).

A manager determining firm boundaries must take this effect into account. Theoretical approaches emphasizing the firm as a repository of routines (Kogut & Zander, 1996; Nelson & Winter, 1982), must also consider that these informal elements put limits on an organization’s ability to acquire external information and adapt to changing circumstances (Leonard-Barton, 1995). Faced with these two opposing effects, managers will need to monitor the development of firm-specific routines and, if necessary, expose internal units to external sources of information by severing existing boundaries.

The discussion above shows that informal institutions within firms are fundamental in determining the limits of firms. While certain informal institutions solve market failure within hierarchies, they also bring their own costs. Thus, informal channels improve coordination but creates avenues for influence activities; social norms create trust and facilitate cooperation but induce biased decision-making and trigger social comparison processes; firm-specific routines are instrumental in facilitating the internal exchange of information but lock the firm out of external sources of information. The paradox is that the same informal processes that help to mitigate market failure create hierarchical failure once they are internalized within firms. We stress that a careful examination of such processes is necessary to develop a more complete theory of the firm—a theory not only based on how firms expand their boundaries in the
presence of market failure, but also how they sever their boundaries in the presence of hierarchical failure.

CONCLUDING REMARKS

NIE has contributed to our understanding of how formal institutions such as explicit incentives, authority, contracts, and ownership can be aligned to exchange conditions to increase organizational performance. Organization theory and economic sociology, in turn, have provided a deep examination of the role of informal institutions such as social norms, trust, routines, and political processes. However, by focusing on each type of institution in isolation, scholars have not paid sufficient attention to the interplay between formal and informal institutions and its performance implications. It is our goal in this paper to demonstrate how the joint assessment of the role of formal and informal mechanisms governing exchanges within and between firms provides new insights and expand the explanatory power of existing theories of organization.

The insights generated by this line of research have important implications for business strategy. It is widely recognized that informal institutions have a critical role in the performance of organizations. Informal institutions can be either performance enhancing, such as relational governance, or performance damaging, such as influence activities. In some cases, the same informal institution can promote or undermine performance depending on the circumstances; for instance, firm-specific routines can improve coordination but at the same time reduce an organization’s ability to respond to external changes. Thus, it is fundamental that scholars not only outline the benefits or drawbacks of informal mechanisms, but also inform managers about how to adjust them. In most cases, managers have only formal mechanisms at their disposal to
change the trajectory of informal processes. Thus far, theories of organization have provided insufficient guidance on how informal institutions can be shaped through changes in formal institutions or by other means.

The good news is that there is still much research needed in understanding the interaction between formal and informal institutions and exploring how this articulation can potentially deliver superior performance for those who manage institutions. A more extensive collaboration among disciplines that have traditionally focused on each type of institution will certainly be a necessary step in developing this stream of research.
REFERENCES


