No. 2006–34

BEYOND PENROSE: A COGNITIVE THEORY OF THE FIRM

By Bart Nooteboom

April 2006

ISSN 0924-7815
Beyond Penrose: A cognitive theory of the firm

Bart Nooteboom
Tilburg University
b.nootooboom@uvt.nl
mail address: Adelheidstraat 82, 2595 EE the Hague, Netherlands
tel. +31 703478605

April 2006

Abstract

This paper uses a cognitive theory of firms and organizations, with a focus on learning and innovation. Here, cognition is a wide notion, including value judgments and corresponding feelings and emotions. This paper focuses on the relation between that cognitive theory and Penrose’s theory of the growth of the firm. As in Penrose’s work, the focus is on learning, rather than on efficient utilization of resources or appropriation of returns from them. Also as in Penrose, the underlying view of cognition is a constructivist one, according to which people with different experience view the world differently. So far, the paper is consistent with Penrose. However, it also adopts and further develops some of the criticism of her views, concerning the role of other human resources than managers in organizational learning, problems of conflicts of interest and governance within the firm, dynamic capabilities for developing new capabilities, and, above all, the alternative of collaboration between firms, for learning and innovation, in the combination of capabilities between rather than within the firm. In particular, it argues that, in contrast with Penrose, there are limits to firm size.

JEL code: D01, D21, L22, M13, M14, O31

Keywords: theory of the firm, Penrose, knowledge, learning, innovation, dynamic capabilities, firm size, growth of the firm.

The central question of this paper is: what are the sources of innovation? Beyond the efficient utilization of resources, and appropriability of their returns, it focuses on their creation. While this may seem new from the perspective of recent economic theory, which has focused on efficient or profitable use of existing resources, some scholars have shown that the question of resource creation goes back as far as Adam Smith, for whom division of labour and new resource creation went together. Best (2002: 179) quoted Smith as connecting division of labour with ‘new improvements of art’.

Penrose (1959) suggests that the sources of innovation lie in firms (Pitelis 2002). However, while Penrose’s account of the growth of the firm includes managerial learning, in the discovery and utilization of as yet unutilized potential of existing resources, and suggests that this moves on to the development or adoption of new resources, it hardly shows how the latter is done. In other words, dynamic capabilities are assumed rather than analysed. To use the terminology of exploitation and exploration
elements to be combined come from, and where does the combination occur? My answer will be as follows. The elements for novelty come from markets (knowledge indeed being dispersed, as Hayek claimed), firms serve to provide a niche for entrepreneurial vision (as Schumpeter proposed) and to carry it into realization, in ongoing novelty of combinations of potential services offered by resources (as Penrose proposed). Here, I am bringing together Hayek, Schumpeter and Penrose, one could say. They all share a notion of differential cognition, i.e. different people having different perceptions, views, and understandings, and that is also a cornerstone in my approach.

The basic idea behind my theory is the Schumpeterian view that a firm serves to establish and implement a particular cognitive focus, setting it apart from the variety of views outside the firm. This view can also be seen as going back to Marshall, who saw the firm as a form of organization that manages and develops knowledge (Loasby 2002, Foss 2002, Richardson 2002). However, cognitive focus causes myopia, and while it enables the implementation of a novel view, it also limits the innovative potential of the firm in novel views. Innovation requires a view of novelty plus the ability to implement it (Ghoshal et al. 2002). A view requires implementation but implementation requires a limitation of view. To repair for this myopia, firms need complementary cognition from relations with outside firms with a different cognitive focus (Nooteboom 1992). In this way, next to learning, innovation and entrepreneurship, inter-firm collaboration also forms an integral part of my theory of the firm.

This paper can only present part of a cognitive theory of the firm, and it focuses on the relations of this theory with that of Penrose. It builds on Penrose, and goes beyond her view of the firm. In particular, it extends learning beyond management, analyzes in more detail how capabilities are combined and developed, extends the analysis of the limits to size and growth of the firm, and includes inter-firm collaboration as an integral part of the theory of the firm, showing how it complements the firm. While Penrose concludes that there are no limits to size, I argue that there are. The limits to the growth rate of the firm resulting from the present analysis are similar to those suggested by Penrose, but add some to it.

In a first section, the paper summarizes the views of Penrose, some points of criticism, and resulting issues to be discussed in this paper. Next, it sets out the arguments
for the need for organizational cognitive focus, specifies features of such focus, and the resulting role for organizations. In a third section, it discusses the origin and stability of cognitive focus, different levels of organization, and limits to the size and growth of the firm, in a trade-off between integrating capabilities in a firm and connecting capabilities between firms.

**Penrose and dynamic capabilities**

Penrose (1959) proposed that firms achieve competitive advantage on the basis of organization-specific resources. It is not the resources themselves that yield results but the services that they may render. As they employ the firm’s resources, managers discover new ways of employing them, in novel combinations, in response to entrepreneurial views of opportunities, and this provides a basis for ongoing growth of the firm. Such entrepreneurial views are cognitive constructions that are unlike objective reflections of reality, vary between people, and are therefore idiosyncratic. This view goes back to Boulding (1956), among others. The present article similarly adopts a cognitive constructivist view, yielding variety of cognition between people. People construct their cognitive categories of perception, understanding and evaluation in interaction with other people. As a result, people view the world differently, yielding ‘cognitive distance’ to the extent that they have developed their cognition along different life trajectories (Nooteboom 1992, 2000).

Penrose (1959) proposed that the size of the firm is not constrained. Firm size is not constrained by limits to economies of scale, or diseconomies of scale, related to products or the size of their markets, since firms can expand by adding new products to their portfolio. Nor are there diseconomies of scale in management. Firm resources are never completely utilized, and yield scope for further extension of activities and capabilities. Penrose proposed that the rate of growth of the firm is constrained by the scope of managerial resources, in particular the ability of existing management to select and introduce additional management and the rate at which such incoming management can adapt to existing plans, procedures etc. She noted that diversification is limited by the need to maintain necessary integration with the rest of the firm, and avoid bureaucracy (op cit.: 208), and that thus there is a crucial trade-off between speed of expansion and maintenance of control (op cit: 189), in a ‘fundamental ratio of managerial resources available for expansion’.

However, Penrose neglected other problems involved in continued expansion of resources and capabilities, internally or by merger or acquisition, the need to also divest, the need to maintain focus, and alternative opportunities of growing by collaboration with other firms rather than by expansion, as has been widely recognized in the literature (Pitelis 2002). Corresponding with this, Penrose had too rosy a view of the capabilities of large firms, and neglected the potential of smaller firms. While Penrose’s account fitted well with the development of capitalist firms in her day, since then there is much evidence of de-conglomeration, downsizing, divestment and sharpening of focus. Kay (2002) documents how the Hercules company, which was a central source of inspiration for Penrose, in its later development ran into failed diversification and had to divest and to shift its core. Cantwell (2002) showed the need for coherence in the technological and
productive activities of the firm to continue to innovate. Lazonick (2002) showed how after the wave of conglomerations in the nineteen-sixties and early seventies, from the nineteen-eighties large corporations reduced their range of activities, and how in the ‘new economy’ firms focused on concentrated skill bases. Patel and Pavitt (1997, 1999, 2000) showed, on the basis of technological profiles constructed from patent data, that while firms indeed incorporate a considerable scope of technological areas, as predicted by Penrosian theory, the profiles of firms are remarkably stable, indicating limited changes of composition outside a given focus of technological areas. Thus, the empirical evidence indicates that while there may indeed be a wide scope for combining complementary capabilities, scope, and hence the size of the firm, is subject to limits.

Therefore, in contrast with Penrose I will argue that from the perspective of organization as a cognitive focusing device, and from a perspective of dynamic capability, there are limits to firm size. As a firm grows by adding new activities, at some point it will have to add new capabilities, and as it continues to do so it will dilute its focus too much, slow down its rate of innovation and reduce its flexibility for novel configurations of capabilities, compared to opportunities for engaging in more variable and more exploratory patterns of collaboration with other organizations, in alliances and networks.

A second well-known point of criticism of Penrose that I share concerns the exclusive role that she accorded to management, in the identification of opportunities and in the learning of new uses of resources (Pitelis 2002). Since the time when Penrose wrote her book, here also considerable change has occurred, in growing skepticism as to how managers would be able to identify new opportunities better than staff that actually operate technology, and interact with customers and suppliers. Nowadays, we are more inclined to accord innovative potential to human resources more generally. I propose that there is a corresponding shift in the task of management. The central task no longer lies in identifying opportunities and guiding novel combinations, but, I propose here, on a meta-level of managing cognitive focus in order to enable people to understand each other and collaborate with each other, in their identification and implementation of opportunities, and to set cognitive focus in answer to the question how to combine exploitation and exploration, within or between organizations. In other words, managerial resources are seen as lying primarily in guiding and coordinating cognition in the firm.

I accept Penrose’s view concerning the limit to the growth rate of the firm, with the difference that I focus on human resources more widely, discounting managerial ability to know and foresee all, so that the constraint becomes that of incorporating new staff more generally. The question is what, more precisely, it is that takes time for incoming staff to adapt to. Penrose (op. cit: 206) proposed that the growth of total supply of management services is faster than growth of the firm, up to a point, and then possibly declines. Why would that be? The view given in the present paper is similar to that of Penrose, but more specific, in that the firm, and organizations more widely, are seen as limited by the ability to coordinate cognition in the firm, and new entrants to the firm, whether managers or other staff, need to adapt to what I call the cognitive focus of the firm.

A third well-known point of criticism of Penrose that I share is the neglect of internal conflict of interests and views, and corresponding problems of authority, monitoring, control, incentives and motives. In other words, a theory of the firm should include not
only issues of competence but also issues of governance. In my theory, that is included in organizational focus, which has a competence side, for mutual understanding, and a governance side, for ability and willingness to collaborate.

With her view of organization-specific resources, Penrose inspired a stream of ‘resource’, ‘competence’ or ‘capability’ based theories of the firm, in the management and organization literature. In the literature, that view is claimed to stand in contrast with the ‘market positioning view’ attributed to Porter (1980, 1985), derived from industrial organization economics.

However, it has been claimed, e.g. by Foss (2002) and Goshal et al. (2002), that most of the resource/competence/capability literature focuses not on learning and the creation of new resources but on the utilization of resources once they are created, in particular on appropriability by some ‘isolating mechanisms’ from competitors, and thereby is closer to traditional industrial organization economics and Porterian views than they make out, while ‘Penrose stresses entrepreneurship and learning in world characterised by change and uncertainty’ (Foss 2002: 156). Also, that literature did not implement Penrose’s view of cognitive differentiation between people. Hence, since this paper focuses on resource creation, innovation and learning, and cognitive differentiation, I will not include most of the literature based on the Resource or Competence based View. However, here and in the further development of a cognitive theory of the firm, beyond this paper, I do include the ‘dynamic capability’ view that developed later (Teece et al. 1997, Dosi et al. 2000), and that did focus on learning and innovation. In particular, I include a branch of the capability view is that of the ‘knowledge based’ theory of the firm that emphasizes capabilities in the form of knowledge (Kogut and Zander 1992, Quinn 1992, Nooteboom 1992, Zollo and Winter 2002).

Teece et al. (2000: 339) proposed that ‘Dynamic capabilities .. reflect an organization’s ability to achieve new and innovative forms of competitive advantage despite path dependencies and core rigidities in the firm’s organizational and technical processes’. So far, the literature has offered limited insight in how that is done, and this forms the focus of my cognitive theory of the firm. Part of that is included in this paper.

Organizational cognitive focus

Using elements from definitions from McKelvey (1982) and Aldrich (1999), I define organizations is as follows:

*Organizations are myopically goal-directed, socially constructed, more or less focused systems of coordinated activities or capabilities.*

The notion of organizational goals is problematic (Scott 1992). When are goals personal and when collective? Are goals those of an entrepreneur, of managers, or of some ‘dominant coalition’ (Cyert and March 1963)? Avoiding that issue, in the present paper, I assume that organizations do have goals, emerging from guidance by entrepreneurs, response and initiative from others, in multiple interactions, within and between firms. Personal goals are different from organizational ones, but they may be,
and for viable organizational membership should be, aligned, in part, with organizational goals.

Organizational cognition also has its problems. Cognition as mental activity by definition cannot apply to aggregates such as firms or organizations. However, such aggregates can be seen as engaging in the use and production of knowledge, and people in an organization can share views, interpretations, understandings, values and norms of behaviour, which are not shared outside the organization, in organizations as ‘systems of shared meanings’ or ‘interpretation systems’ (Smircich 1983, Schein 1985, Weick and Roberts 1993, Weick 1995, Cook and Yanow 1996). Here, I use the notion of organizational cognitive focus (Nooteboom 1992, 1999). Guided and constrained by organizational focus, people in organizations may accept organizational goals, and contribute to their elaboration, maintenance, drift or shift, as an outcome of interaction between them, and between the organization and its environment. Initially, at their founding, firms carry a strong imprint, in both goals and cognitive focus, from the founding entrepreneur(s). Subsequently, both goals and focus are subject to reconstruction and change as the outcome of interactions in the firm. This is indicated by the term ‘socially constructed’ in the definition of organizations. The ‘more or less focused’ nature of these systems refers to the condition that organizational focus can be more or less encompassing and cohesive. The features of organizational focus are elaborated later.

Goal direction of organizations is ‘myopic’ for several reasons. First, individual cognition is ‘bounded’ in that it is contingent upon mental categories that both enable and constrain cognition, and arise from experience along specific life trajectories. In this ‘activity based’, ‘embodied’ cognition, insights, preferences and goals do guide action, but they are also constructed from it, and then follow rather than proceed action (Weick 1979). This makes cognition bounded not only in the sense that one has a limited capacity for rational evaluation, but in the more fundamental sense that one’s perspective is biased by experience and subject to unforeseeable development. In organizations, individual cognition is further guided and constrained by organizational focus. And finally, shifts of goals and focus of organizations are emergent, and hence unforeseeable, outcomes of complex interactions between people. In this process there arise discrepancies between official ‘espoused’ organizational goals, and actual goals ‘in use’ (Argyris and Schön 1974).

There is a range of possible goals, such as survival, profit, social or political legitimation, competitive strength, political influence, creativity, freedom or independence. Achievement of goals requires coordination of actions, using means according to relevant know-how, under external commercial and institutional conditions. Know-how includes individual-level and group-level capabilities, in skills and knowledge that may be technical, organizational and behavioural, and organization-level capabilities to configure and coordinate them. There is, in general, a multitude of ways to select and configure goals, actions, means, know-how and external conditions. Coordination of specific activities is not necessarily only a task of management, and may arise on other levels, by plan and design or more spontaneously. Coordination may be achieved by assigning capabilities to jobs, assigning people to jobs, thereby defining the roles they have in the organization, and defining and governing relations between them. Here, Mintzberg (1983) proposed five forms of coordination: specification of outputs,

In these configurations of goals, actions, means, know-how and external conditions, choice is needed. One cannot strive for everything at the same time. One cannot look in all directions at the same time, and if it were possible, one would probably see nothing. In other words, there is a need for focus on goals, actors and actions, means and know-how, and on how to configure and coordinate them under what conditions. The ability to do that constitutes organizational capability. The central difference between firm and market is that in the former such focus is made and in the latter it is not, or to a much lesser extent (there still is a remaining, shared cognitive focus from shared culture). Thus the market has the higher potentiality of variety of performance, and the firm has the higher actuality of performance.

The Need for Focus

Focus means limitation of a range, of activity, ownership, attention, meaning or capability. Then, organizational focus can mean limitation of its range of activities, in terms of products, markets and technologies, of physical, cognitive or cultural assets, of individual or organizational capabilities, or a combination of all of those.

Many firms still define themselves in terms of specific activities, but as the pace of change of knowledge, technologies and markets increases, firms are learning to shift their focus from a given range of activities, products, production processes, physical assets, distribution channels and the like, to a focus of underlying capabilities that have the potential of generating and supporting a variety of products and production processes (Quinn 1982, 1992). This yields greater flexibility to adapt to changing markets and technologies with changing product lines and production processes. That was also one of the essential points made by Penrose (1959). Capabilities (resources) can be applied to a variety of specific activities (‘services’).

Next, and beyond Penrose, there is increasing pressure to narrow the focus of capabilities to those ‘core capabilities’ in which firms can maintain durable competitive advantage. Such focus yields a further increase of flexibility, in a wider scope for more variable configurations of the firm’s capabilities with those of other firms. It also improves their distinction with respect to other firms, with the usual consideration, from industrial economics, that differentiation yields higher profitability than price competition with highly substitutable products. I will argue that such focus also increases speed of innovation, which is vital under increasing competition in all dimensions, of price, quality and innovation. It also increases cognitive diversity, in collaboration between firms, as a source of innovation (Nooteboom 1992). Also, the external configuration of capabilities between firms is enabled by new technical opportunities, from information and communication technology, and emerging organizational capabilities, in the coordination of activities between firms. I will argue that, in contrast with Penrose (1959), all this yields limits to the size of firms.

For any given focus of capabilities, in order to profit from complementarities between them cognitive coordination is needed on their interfaces, in the wide sense of cognition used here, to enable sufficient mutual understanding and ability to collaborate (competence), and willingness and commitment to do so (governance). Such alignment of
cognition requires what I call ‘organizational cognitive focus’. I propose that this constitutes a central organizational capability.

Note that in cognitive coordination not everyone has to have the same ideas on everything. In fact, it may even be that there is not a single idea shared by all. Diversity of ideas is good for innovation, and autonomy and room for initiative are often good for motivation. So, one should aim for minimum alignment of cognition needed to utilize opportunities from complementary capabilities. Variety of cognition should be limited only when needed for the feasibility and efficiency of collaboration.

On the competence side, focus is needed to enable people to understand each other and connect complementary knowledge, without unduly restricting variety and creativity. On the governance side, focus is needed to motivate people to collaborate and share and connect knowledge, without unduly restricting autonomy, ambition and competitive spirit. Next to coordination, organizational focus also has functions of selection and adaptation. In selection, it selects people, in recruitment and often on the basis of self-selection of personnel joining the organization because they feel affinity with it. In adaptation, it socializes incoming personnel, with initiation, and focuses their capabilities, in training.

**Features of Focus**

The question is how far organizational focus goes, or should go. Both inside and outside organizations, people have more goals, capabilities, roles and relations than those that are governed by organizational focus (Dimaggio 1997). Ring and van de Ven (1994) made a distinction between organizational roles people play and their behaviour ‘qua persona’. This was presaged by the distinction Simmel (1950[1917]) made between a person’s function in an organization, which takes up only part of his personality, and his full personality. So, one question is how far organizational focus reaches in affecting actions of people. Berger and Luckmann (1966) distinguished between primary socialization in family, as one grows up, and, building on that and molding it further, secondary socialization in places of work.

The content and extent of cognitive alignment in in organizations may vary. In addition to the distinction between the competence and governance sides of focus, there are five dimensions for both. First, there is width, i.e. the range of different areas of competence and governance in a firm to which focus applies. This depends on the range of capabilities that a firm encompasses. Second, there is reach, i.e. the number of aspects within each area covered by the focus. Does it affect all or only some key aspects of a given capability? A third dimension is tightness versus looseness, i.e. narrowness of tolerance levels of standards or rules imposed by focus, versus allowance for slack and ambiguity, with improvised, unforeseen meanings, actions, etc.

Fourth, focus may have different content. In particular, on the governance side it may be formal, i.e. depersonalized, norms of legitimacy, which regulate what managers and workers can legitimately do and can expect from each other. Such norms render relations more impersonal and thereby reduce tensions associated with the exercise of personal power, and they enlist workers to participate in the control of their colleagues (Scott 1992: 306). The content of focus may also be more cultural, in the sense of offering guidance by more emotion-laden underlying values, expressed in symbolic entities,
behaviours, events or processes. The two types of content are related, since norms of legitimacy may be expressed culturally, but can nevertheless be distinguished. One can have norms of legitimacy that are specified rigorously and formally, and one can have more informal, ambiguous, cultural features that go beyond norms of legitimacy.

Fifth, focus may relate to surface regulations concerning specific actions or to underlying more fundamental notions, in a deep structure of logic, principles or cognitive categories that form the basis for surface regulation. Simon (1976) already acknowledged that an organization controls not decisions but their premises. Nelson and Winter (1982) made a similar distinction, between routines and ‘meta-routines’ that guide the development of routines. As already indicated, Mintzberg (1989) allowed for ‘missionary’ organizations. Schein (1985) made a similar distinction in organizational culture. Below surface features such as specific rules, practices, symbols, myths, rituals, at the basis of organizational culture lie fundamental views and intuitions regarding the relation between the firm and its environment (‘locus of control’: is the firm master or victim of its environment), attitude to risk, the nature of knowledge (objective or constructed), the nature of man (loyal and trustworthy/self-interested or opportunistic), the position of man (individualistic or part of a community), and relations between people (rivalrous or collaborative), which inform content and process of strategy, organizational structure, and styles of decision-making and coordination. Schein also allowed for an intermediate level, connecting the fundamental cognitive categories with the surface level of specific structures and rules, in the form of general principles that express fundamental cognitive categories but are yet general and generic rather than specific to certain activities and contexts.

The difference between activities, surface regulation and deep structure is schematically illustrated in Figure 1. Here, for simplicity of exposition, the intermediate level is left out. A given surface regulation enables a bundle of potential actions. An underlying cognitive category in deep level structure enables a bundle of surface level regulation. The establishment of coordination on the surface level (routines, if one wants to use that term) leaves freedom for variety of underlying cognitive categories, but has to be set up ad hoc each time, and requires the solution of complications due to differences in underlying cognition. The establishment of coordination on the deep level yields more ex ante agreement for setting up surface regulation, and thus enhances speed of action, but it reduces variety of cognition on the deep level. It entails more indoctrination. I will argue that organizations serve to coordinate on the deep level, with an advantage of easier and faster understanding and agreement, while collaboration between organizations operates more on the surface level, with the advantage of greater variety on the deep level.

Nooteboom (1999) employed the notion of ‘cognitive distance’, i.e. differences in cognition between people. The concept entails a distinction between reducing and crossing cognitive distance. Reducing cognitive distance entails alignment on the deep level of cognition, so that people think more similarly. Crossing cognitive distance is
making surface agreements while maintaining differences on the deep level, with people continuing to think differently. When people who think differently continue interaction, starting from surface agreements, they may in time come to think more similarly, i.e. share underlying cognition, in a reduction of cognitive distance.

**Organizations**

If capabilities are connected between A and B, and between B and C, and not between A and C, there is a need for A and B to have some cognition in common, and for B and C, but there may be no need for A and C to have anything in common. This situation may arise in sequential interdependence (Thompson 1967). Under pooled interdependence (ibid.), B, C and D may all share a common resource A, as a resource they tap from, or a resource they contribute to, in a ‘star’ or ‘hub and spokes’ configuration. Here they must all have some cognition in common with A, but not necessarily with each other.

Why would such arrangements have to take place within an organization? This is the classic question, ever since transaction cost economics, of why the firm exists as an organization that integrates different activities or capabilities. Why not have autonomous, independent capabilities, and set up sufficient mutual understanding and agreement between them ad hoc, only when the need arises?

One reason for integration in a firm is technical complexity and systemic coherence. A system of production is systemic or complex to the extent that there are (1) many components that are (2) densely connected in ties of complementarity, which are strong in that connections concern a wide range (3) of actions with (4) narrow tolerances on interfaces in order to preserve systemic coherence. As a consequence, in a complex system a small local change, in some component, may trigger multiple changes elsewhere, reverberating through the system, to yield a very different configuration.

This well-known principle from systems theory has been revived in recent literature (Teece 1986, Langlois and Robertson 1995, Levinthal 2000, Postrel 2002) because of its importance for the stability of organizational systems and for differences between firms within an industry.

More systemic coherence requires more cohesiveness of focus. A greater number of components requires greater width of focus, a wider range of interaction requires a greater reach of focus, and narrower tolerances require greater strictness of focus. A less cohesive focus, in a de-coupled system, may yield loss of efficiency, but has dynamic advantages. It facilitates the outsourcing of components and the keeping of only components that include the organization’s core capabilities and complementary resources that are inseparable from them. It increases the scope and variety of possible linkages, and greater flexibility in the configuration of activities or capabilities. One form of decoupling is modularization, i.e. by re-composing the system in components with a one-to-one correspondence with function and/or capability, and with standards on the interfaces between component activities (cf. Langlois and Robertson 1995).

A second point, to be considered here, is derived from the notion of specific investments from TCE (transaction cost economics). If the set-up of sufficient cognitive alignment constitutes a specific investment, which cannot be used in other linkages, it is only efficient to make it under the prospect of sufficient duration or intensity of
utilization to recoup the specific investment. Especially when the knowledge involved is tacit, the building of understanding takes time, in shared practice, and it is likely to be fairly specific to the linkage involved. This has an implication for the durability of linkages, but not yet for a durable combination of different capabilities within an organization. Collaborative relationships that are sufficiently durable to build up and utilize mutual understanding may also be achieved between rather than only within organizations (Nooteboom 1999).

A third, more crucial, cognitive, argument for integration of capabilities within an organization derives from Schumpeter. Novel entrepreneurial ideas typically do not make sense in an established institutional order of dominant ideas, routines, recipes, logics and practices, in an industry (Spender 1989, Bettis and Prahalad 1995). Here, organization yields a sheltered niche for deviance of ideas that cannot be traded or connected outside, due to lack of comprehension, and need to be configured within the organization, on the basis of entrepreneurial vision and charisma to inspire followers.

A fourth point for organization, which applies also to firms with an established, recognized position, is the following. The whole point of moving from specific activities to more generic capabilities, derived from Penrose, and indicated above, was to obtain more scope for novel activities, in novel uses of capabilities, and thereby have more flexibility to respond to increasingly rapid changes of threat and opportunity, in technology and markets. Now novel utilization of capabilities and the new configurations of capabilities to which this is likely to lead, requires renewed investment in mutual understanding and ability to collaborate, if that was not already in place. That, however, takes time, particularly when the knowledge involved is more tacit. As a result, to maintain scope and flexibility for novel utilization, and hence novel combinations, of capabilities, it pays to have a further reaching, more cohesive scope of mutual understanding, and ability and willingness to collaborate, already in place, beyond linkages that are currently operational. This is an argument of speed and flexibility. To warrant and utilize such wider sharing of cognition, across not only actual but also potential, unpredictable new linkages, plus the need for a certain duration to achieve and utilize the investment involved, yields an argument for organization, with a range of capabilities that are more tightly and durably aligned by organizational cognitive focus. For the more ad hoc investment in linkages as the need or opportunity arises, organizations can employ inter-organizational alliances.

The cognitive argument for organization can be further sharpened with an analysis of the depth of cognitive focus. For the same reason that a focus on capabilities yields more scope and flexibility of activities than a focus on activities, cognitive alignment in deep structure rather than in surface regulation is more generative and flexible, yielding a wider scope of capabilities. This was illustrated in Figure 1. To use a metaphor, fundamental mental categories are like the roots of a tree, largely invisible in the subconscious, capabilities are like the branches, and activities are like leaves and blossoms that appear, die off, and re-appear. Deep level cohesiveness is more tacit than surface level rules, and hence requires more time to develop and is more specific, requiring longer time to recoup as an investment, and therefore is more a feature of organizations than of inter-organizational relationships that, on the whole, tend to be less durable, less cohesive and more superficially coordinated.
There is another reason for coordination on a deep level, especially on the governance side. That lies in the condition that work has become increasingly knowledge intensive, professional, and abstract, and thereby more difficult for management to monitor and evaluate, let alone measure. This makes rule- and contract based control more difficult, and creates an advantage for more intrinsic motivation that requires less monitoring and control, on the basis of underlying values of conduct, or bonds of empathy, or even identification, and routinization, that require more time to develop and more cohesion to function (Nooteboom 2002).

A more cohesive focus may also cater to the social need of people to have a sense of belonging. It does, however, entail more ideology and indoctrination, which also has its downside, in organizational myopia, group think, lack of scope for creativity, and lack of freedom, possibly subjugation, or even a form of serfdom.

In sum, the most fundamental function of organizations, in contrast with inter-organizational relations, is to provide a focus with a certain depth, for the sake of coordination that can quickly shift to novel patterns of activity, and for motivation that is more intrinsic. A cognitively and culturally more cohesive group, within an organization, can more quickly alter patterns of collaboration that lie within its potential. Less cohesive inter-firm collaboration, with more cognitive variety, has a wider scope of potential novelty, but requires more time, in setting up surface regulation, to utilize opportunities. I propose, and will later argue in more detail, that this consideration yields limits to firm size and yields boundaries of the firm. I will argue that as the range of capabilities to be coordinated increases, the resulting increase in width of focus entails an increase of reach and depth of focus, and that this increase of cohesiveness increasingly limits the cognitive variety and flexibility needed for ongoing innovation by novel combinations.

**Development and Stability of Focus**

As noted earlier, organizational focus emerges from the imprint of the entrepreneur who started the organization, is subject to some drift due to interaction between staff, turnover of staff, and due to shifts resulting from crises, caused, in particular, by shifts in the environment, or by new, challenging interpretations of the environment. Also, of course, its existence depends on population effects, in the weeding out by competitive selection. When resources are scarce and competition is tight, selection is likely, in the long run, to yield organizational cognitions and structures that reflect the exigencies of the environment of markets and institutions.

An important condition for the features of focus is whether the organization engages in exploitation, exploration, or both. This depends on competitive opportunities and pressures, as a function of innovative and market turbulence in the environment, and on strategic choice. For exploration cognitive distance needs to be larger than for exploitation, with a focus with relatively little reach and strictness.

Focus also depends on outside legal and cultural institutional conditions. In particular, what a firm can and needs to do in governance depends on contractual opportunities, norms and values of conduct, intermediaries of many kinds, political conditions, etc.

Organizational focus cannot be integrally and instantly re-shaped as a function of experience in competition, and to a greater or lesser extent this yields organizational stability or even inertia (Hannan and Freeman 1989). Such stability is a well-established
empirical phenomenon. In their study of large firms Patel and Pavitt (2000: 317) found that: ‘90% of firms have profiles of technological competence that are statistically similar between 1969-74 and 1985-90’ which ‘remains true even after taking account of acquisitions and divestments’, and ‘. . . of 41 of the largest firms only one had a technological profile statistically different’.

There are several reasons for stability of organizational identity and focus. One reason is systemic. As indicated before, in a complex system, with many tight linkages between components, change in any element may endanger systemic integrity, and would then require multiple changes elsewhere in the system, to arrive at a new feasible configuration.

A second reason is that cognitive focus yields absorptive capacity (Cohen and Levinthal 1990) that tends to mostly confirm itself in its functioning (imprinting). Ways of seeing and making inferences yield habits that are relegated to subsidiary awareness, in routinization. However, while there is inertia, firms may be able to escape from it, on the basis of appropriate dynamic capabilities. However, that subject cannot be discussed within the confines of the present paper.

Cultural identity and cognitive focus are maintained, in spite of turnover and exchange of staff, because in the entry into an organization there is, as noted before, self-selection according to expected fit to organizational culture, as well as adaptation by socialization into organizational culture, in introductory courses, meetings or rituals, and ostracism of those who do not conform. Furthermore, according to the idea of intelligence as internalized action the further development of cognition reflects the environment, in this case the organization, in which it takes place.

Levels of Organization

Generally, organizational focus will be narrowest in single-person, owner-manager firms, wider in work groups, wider yet in larger firms consisting of multiple groups, and widest in multi-divisional firms, as a function of the diversity of capabilities involved.

In a single-person owner-manager firm, by definition cognitive distance is zero. There, organizational focus is inclusive. Personal life and business are highly interwoven. This arises in finance, with personal or family capital, profits as a source of personal income, and business risk as a personal risk. It often arises in housing, with the firm located at the home. It also arises in the goals of the firm. Organizational goals are more varied and more personal in small firms. Self-employed people may flee from the authority imposed by employers, they may take refuge in self-employment from unemployment due to economic crisis combined with lack of social employment benefits, or discrimination of race, ethnicity or gender in labour markets. They may seek independence as a goal in itself or to achieve the opportunity to implement deviant ideas, satisfy personal preferences concerning niche products, or small-scale production with informal relationships and absence of bureaucracy.

The small firm often has a limited portfolio of capabilities, yielding a limited range of technologies, products and competencies. As a result, small firms are vulnerable, with limited diversification of risks, limited specialization in functions, limited economies of scale and scope, and limited career perspectives for personnel. They also have both the potential advantages and disadvantages of a cohesive focus, in thick and tight, often highly
personalized relationships, with limited division of labour. High cohesiveness may also result from the cognitive stamp that the entrepreneur puts on his small organization, where he interacts directly with his personnel. In this way, radically innovative, small firms may isolate themselves and thereby close themselves off from the sources of application and further innovation.

This yields one of several paradoxes of the small firm. On the one hand, small size, with personalized, thick, informal relationships, integration of tasks among few people, and direct contacts, internally and outside, e.g. with customers, enables high flexibility and motivational power of identification with the firm. On the other hand there is potential for suppression of freedom and variety, and of isolation from the environment (cf. Nooteboom 1994).

In larger firms, especially distance in job-related competence is larger, in a wider and deeper division of labor, but distance on the moral side of cognition is not necessarily larger than in a small firm. In fact, since in face-to-face work groups there is more informal, spontaneous social control of free-ridership (Simmel 1950[1917]), in the small firm the need for a more explicit moral focus is less. In larger organizations more attention may be needed to the moral dimension of organizational focus across different work communities. A difference in culture between large and small firms lies in the fact that with a more extensive division of labor, with coordination between greater numbers of people across possibly distant organizational units, knowledge and rules need to be codified to a greater extent than in small firms, where coordination can take place by direct supervision (Mintzberg 1983).

Limits to Size and Growth of the Firm

Are there limits to the size of a firm? Penrose (1959) claimed that there are none. Firms can grow by diversification and by employing new services that belong to the potential of existing resources and their configurations. However, in ongoing growth at some point new capabilities have to be added. I expect that there are limits to size because under increase of the range of capabilities, the organization faces a fundamental trade-off between variety and coordination. Either variety is maintained at the expense of coordination, and then the question arises why the elements should be part of single firm rather than being independent, or coordination is maintained at the expense of variety, which reduces innovative potential. I now proceed to analyze the issue in more detail.

If the number of capabilities, and resources more generally, is n, then the number of possible connections between them is n(n-1)/2, and thus increases quadratically in n, if everything remains connected with everything else, yielding an accelerating increase of costs, including opportunity costs, of coordination, and, probably, noise in communication. This, of course, yields a classic reason for decomposition, in hierarchies or network structures such as, for example, a hub and spokes structure. Capabilities are clustered in local units, such as communities of practice, which are in turn connected in divisions of firms, which are in turn connected in corporations. Is there a limit to this?

From the present cognitive perspective, problems of coordination increase with the dissimilarity of connected capabilities, within the different levels of organization. Also, empirical research by Nooteboom et al. (2005) showed a principle of decreasing returns to
knowledge. As knowledge accumulates, further novelty has to be sought at increasingly greater cognitive distance. As a result, as capabilities accumulate, the dissimilarity of novel capabilities increases, thus increasing problems of collaboration.

The costs of coordination are not just direct costs, but also opportunity costs of loss of cognitive variety, in the fact that in establishing mutual understanding, and willingness to collaborate, across an increasingly heterogeneous batch of capabilities, each of them gets more constrained in its idiosyncrasies of cognition. The paradoxical result then arises that while potential variety increases in terms of the scope of capabilities, in each of them actual variety decreases due to rising needs of common understanding. One alternative, to prevent escalation of costs and complexity of coordination, is to leave potential connections unutilised, but then the question arises why unconnected capabilities should be combined in a single firm.

This brings me close to Richardson’s (1972, 1999) position that activities should be combined within a firm to the extent that they are similar and complementary, and should be relegated to outside relations to the extent that they are not. However, the analysis can be further refined.

A further question is whether coordination should be limited to surface level regulation, in the ad hoc coordination of specific activities when they arise. In that case, the question again is why that would be more efficient inside than between firms. Such ad hoc regulation has to be made anew each time that activities are re-configured. With a lack of deeper cognitive alignment, this takes time, in a working out of differences in cognition, in competence and governance. A faster alternative to ad hoc surface regulation is to build up a store of surface regulations that one might draw from as the need for any them appears. However, it may be difficult to predict what combinations may arise, and for each of them what regulations they might need. With an increasing range of potential combinations their number would increase, and the question arises how much of all that effort will actually be utilized. Thus, for surface regulation among multiple capabilities, the best would generally be to await what concrete activities arise and improvise coordination accordingly.

The advantage of inclusion of a range of capabilities within one organization is that alignment of underlying cognition can be achieved, on a deeper, more generic level of cognition, yielding greater speed, scope and flexibility of generating surface regulations for novel combinations. The (specific) investment in such focus is worthwhile for its perspective, within a firm, for intensive and repeated utilization. Here is a remnant of TCE logic.

However, there are several drawbacks to this. First, it may be difficult to predict which capabilities will yield interesting combinations in an unknown future. Second, as the number of capabilities increases, cognitive alignment across all of them yields reduced variety in each of them, as indicated before. That effect is less to the extent that the capabilities and views are already more similar. Chemistry and biotechnology have more in common than chemistry and information technology. In other words, people can easily coordinate on the surface level to the extent that differences in underlying perceptions, views and convictions remain limited. As underlying differences increase, indoctrinating people with shared perceptions, views and convictions smoothes their collaboration but reduces their variety of cognition.

As a result, it seems that at some point it becomes better not to bring further and more diverging capabilities under a single focus, and to take the alternative of employing inter-
firm collaboration, yielding a wider range of potential capabilities that may yield interesting combinations, and the preservation of more variety in each of them, and to engage in the more ad hoc, time consuming surface regulations for combination when and where the need arises. However, without constructing a shared cognitive focus on the competence side, integration in an organization may still be worthwhile for a shared focus on the governance side, in shared views on how to deal with each other. One can agree on that under great differences on the competence side. But that may also apply to collaboration between firms. Perhaps one can find, with relative ease, partners who are diverse in competence but like-minded on the governance side, yielding a basis for trust.

There is yet another consideration. One may maintain cognitive distance and yet collaborate easily on the basis of a large absorptive capacity, and ability to collaborate more widely. Postrel (2002) asked when communities of practice should invest in knowing about each other, and when they should go their own way. On the face of it, the answer is that they should invest in knowing about each other when activities of different communities are strongly coupled, or, in other words, when activities are ‘systemic’, and that one should go one’s own way when activities are not or only loosely coupled, or ‘stand-alone’. Only in the first case it is necessary to mutually adapt activities. Postrel shows that this intuition is not necessarily correct. If by going their own way, and investing only in their own knowledge and skill, specialist communities can extend the scope and flexibility of their activities, then they can thereby achieve fit to whatever other communities do.

Figure 2 about here

The principle is illustrated in Figure 2, which is derived from Figure 1. Here, an increased absorptive capacity, or ability to collaborate more widely, entails that categories on the deep level of cognition have a widened range of applications, with more opportunities for surface level regulations. With a narrow range, distance between categories of A and B has to be small, as illustrated in the left part of the figure, in order to create overlap for the sake of coordination. With a wide range, on the right part of the figure, with a wider scope of understanding and ability to collaborate, overlap in surface regulation (crossing cognitive distance) is achieved at greater cognitive distance.

In other words, having large absorptive capacity and collaborative ability, and seeking partners with such capacity, would reduce time and effort in achieving requisite mutual understanding. Thus, having high absorptive capacity, and experience in collaboration, and thereby being an attractive partner, becomes a key dynamic capability and competitive advantage. This applies both within and between firms. So, a firm might integrate a great variety of competence and deal with it on the basis of large absorptive capacity and ability to collaborate, in the units that may need to connect with others, with ability to collaborate further supported by an organizational focus only on the governance side.

However, that takes time, in an accumulation of knowledge, for absorptive capacity, and of experience in collaboration with people who think differently, for ability to collaborate. Also, recall that accumulation of knowledge yields the requirement of partners at increasing cognitive distance to learn something new, which may then further contribute to a widening of variety within the firm, with its attendant coordination problems. In other
words: the solution may contribute to a worsening of the problem. Also, having invested in large absorptive capacity and ability to collaborate, one would want to utilize that dynamic capability in a greater variety of different contacts, for which again outside relationships provide more scope. An example here is scholars: the ones who have accumulated most knowledge and most experience in collaborative research engage most in varied collaboration outside their university. In this way, building absorptive capacity within the firm, as a matter of human resource policy, may increase the problem of the most capable staff leaving the firm, before investment in them is recouped.

This rather elaborate analysis shows that it is not easy to refute Penrose. For every problem of firm size, a new potential solution crops up. However, it also appears that every time the end conclusion is that while a solution within the firm may be found, outside collaboration with other firms seems the better option.

In sum, a cognitive limit to firm size appears to lie in a trade-off between coordination and variety of cognition. The advantage of an organization is that by alignment on deeper levels of cognition it affords easier and faster coordination on the surface level of specific combinations of capabilities, but it does so at the price of reducing cognitive variety. In other words, organization tends to improve exploitation at the expense of exploration. The alternative is to engage in slower, more ad hoc, surface level adjustment while preserving cognitive variety, and the potential for that is greatest in outside collaboration with other firms. There may be ways out of this. While maintaining cognitive variety on the competence side organizations may still have a comparative advantage in providing a cognitive focus on the governance side. This applies most in low-trust environments, where indeed organization is most needed for a focus on the governance side. Another option to maintain both cognitive diversity and ease and speed of cognitive coordination is to build and maintain large absorptive capacity and ability to collaborate within the organization. But this requires considerable investment in the accumulation of knowledge and experience, and the question arises whether it would not be more attractive to extend the utilization of such dynamic capability in a greater variety of contacts, outside the organization.

Limits to Growth

To coordinate an increasingly complex whole of capabilities, one needs a sufficiently wide, far reaching and tight cognitive focus, especially on the normative, governance side to maintain ability and willingness to collaborate across such diversity of competence, often across diverse locations, in different cultural settings. This requires a certain strength and sophistication of organizational culture that requires time to develop as well as time for incoming staff to find their way and to adjust. Next to obstacles to size, this yields an obstacle to the speed of growth, which comes close to the limit to the rate of growth identified by Penrose.

Another limit to the rate of growth lies in the building up of absorptive capacity and ability to collaborate, as a dynamic capability to profit from wider cognitive variety, inside and outside the firm.
Conclusion

Penrose’s (1959) theory of the firm, and particularly her underlying view of knowledge as constructed, yielding ‘cognitive distance’ between people, stimulates an analysis of organizational learning and resource creation that goes beyond her own theory. The present paper, utilizing an emerging cognitive theory of the firm, suggests that the fundamental role of organizations is to provide a cognitive focus in order to utilize complementarities between different capabilities. However, there is a trade-off involved with the alternative of profiting from capabilities between different organizations. While within firm coordination on the basis of cognitive focus yields an advantage of easy and fast coordination, it yields a disadvantage of reduced cognitive variety. Conversely, between firm coordination is slower and more laborious, but has the advantage of offering a wider scope of possible combinations. The trade-off point between the two lies in the cognitive distance between the capabilities involved: firms serve for lower distance, and inter-firm collaboration serves for larger distance. Essentially, this confirms Richardson’s view.

An important dynamic capability identified in the analysis is the ability for mutual understanding (absorptive capacity) and ability to collaborate, for crossing cognitive distance, to utilize opportunities for combining different competencies. From the perspective of a cognitive theory of the firm there is much more to be said about dynamic capabilities, but that is beyond the scope of the present paper.
References

Aldrich, Howard

Argyris, Chris and Donald Schön

Best, Michael H.

Bettis, R. A. and C. K. Prahalad

Boulding, Kenneth E.

Cantwell, John

Cohen, Michael D. and Dan A. Levinthal

Cook, Scott and Dvora Yanow

Cyert, Richard M. and James G. March

DiMaggio, Paul J.

Dosi, Giovanni, Richard R. Nelson and Sidney G. Winter (Eds.)

Foss, Nicolai J.

Ghoshal, Sumantra, Martin Hahn and Peter Moran

Hannan, Michael and J. Freeman

Hannan, Michael and John Freeman

Hayek, Friedrich A. von

Kogut, Bruce and Udo Zander

Langlois, Richard N. and Paul L. Robertson

Lazonick, William

Levinthal, Daniel A.

Loasby, Brian

March, James G.

McKelvey, William

Mintzberg, Henry  

Mintzberg, Henry  

Nelson Richard R. and Sidney Winter  

Nooteboom, Bart  

Nooteboom, Bart  

Nooteboom, Bart  

Nooteboom, Bart  

Nooteboom, Bart  

Nooteboom, Bart, Wim P.M. van Haverbeke, Geert M. Duysters, Victor A. Gilsing and Ad van den Oord  
2005  *Optimal cognitive distance and absorptive capacity*, paper under review.

Patel, Pari and Keith Pavitt  

Patel, Pari and Keith Pavitt  
Patel, Pari and Keith Pavitt

Penrose, Edith
1959 *The theory of the growth of the firm*, New York: Wiley

Pitelis, Christos (Ed.)
2002 *The growth of the firm, the legacy of Edith Penrose*. Oxford: Oxford University Press.

Porter, Michael E.

Porter, Miachel E.

Postrel, Steven

Quinn, James B.

Richardson, G.B.

Richardson, G.B.

Richardson, G.B.

Ring, Peter S. and Andrew van de Ven

Schein, Edgar H.
1985 *Organizational culture and leadership*. San Francisco: Jossey-Bass

Schumpeter, Joseph A.
Schumpeter, Joseph A.

Scott, W. Richard

Simmel, Georg

Simon, Herbert A.

Smircich, Linda

Spender, J. C.

Teece, David J.

Teece, David J., Gary Pisano and Amy Shuen

Weick, Karl F.

Weick, Karl F.

Weick, Karl F. and Karlene H. Roberts

Zollo, Maurizio and Sidney G. Winter
Figure 1  Levels of coordination

---

Figure 2  Absorptive and collaborative capacity

---

A  B
narrow collaborative capacity and small cognitive distance

A  B
wide collaborative capacity and large cognitive distance