Beyond Penrose: a cognitive theory of the firm

Bart Nooteboom
Tilburg University
b.nooteboom@uvt.nl

mail address: Adelheidstraat 82, 2595 EE the Hague, Netherlands
tel. +31 703478605

Revised version, August 2006

Abstract

This paper presents elements of a cognitive theory of firms and organizations. Here, cognition is a wide notion, including value judgments and corresponding feelings and emotions. 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. It is argued that the fundamental purpose of a firm is to achieve some minimal cognitive coherence, achieved on the basis of an organizational ‘cognitive focus’. From this, the paper derives boundaries of the firm, and indicates the importance for cognition and learning of linkages between firms. Counter to Penrose it derives an argument that there are limits to firm size.

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

Keywords: theory of the firm, Penrose, knowledge, learning, boundaries of the firm, firm size.

This paper aims to contribute to the development of a Penrosian, cognitive theory of the firm. Penrose focused on managerial learning in developing new strategic images and new services of a firm’s resources. What the present paper adds to Penrosian theory is a wider view of who, other than management, contribute to learning, the inclusion of sources of knowledge outside the firm, and, in relation to that, a view on boundaries of the firm and on the role of inter-firm relationships for innovation.

The roots of a cognitive theory of the firm are as follows. If markets are seldom perfect, and price does not contain all the information needed to know how to operate and compete on markets, then a central task of the firm is to collect data, interpret them into information, transform that into visions, capabilities and knowledge, and to combine those into strategies for markets, products and production. This view may be called Penrosian, but it goes back to Marshall, who also saw the firm as a form of organization that manages and develops knowledge (Richardson 2002: 42). Besides Penrose (1959), that view was further developed e.g. by Richardson (1972, 2002), Kogut and Zander (1992), Quinn (1992), Nooteboom (1992), Fransman (1994), Teece et al. (1994, 1997), Dimaggio (1997), Loasby (1998, 2002), Dosi et al. (2000), Foss (2002), and Zollo and Winter (2002).
Penrose’s cognitive view went beyond information processing to the view that managers produce ‘images’ (a term Penrose derived from Boulding 1956, see Penrose 1959: 11, footnote) and visions. In other words, they create meanings and guide sense-making. This view recognizes the difference between information and knowledge: Different people confronted with the same information will not have the same knowledge, but will develop different cognitive constructions. This view of differential cognition, with different people having different perceptions, views, and understanding, was shared in economics by, among others, Schumpeter (1934), Hayek (1945), Boulding (1956), and (other) Austrian economists (Lachmann 1978). A characteristic problem for an entrepreneur is that his vision cannot be made sense of from the perspective of established practice and ideas. Therefore, one function of a firm is, according to Schumpeter (1934), to shelter novel entrepreneurial vision from the outside world and give it a chance to develop (Ghoshal et al. 2002).

More than the economics literature, on the basis of social constructivist views of cognition the literature on management and organization has further developed the view that firms construct their own, more or less organization-specific meanings and interpretations, in the organization as a system of ‘sense-making’ (Weick 1995), ‘collective mind’ (Weick and Roberts 1993), system of ‘shared meanings’ (Smircich 1983), ‘interpretation system’ (Choo 1998), or a cognitive ‘focusing device’ (Nooteboom 2000). This also relates to the view of entrepreneurs providing ‘cognitive leadership’ (Witt 2005).

Penrose (1959) suggested that the sources of innovation lie in firms (Pitelis 2002a). According to Adam Smith, discovery is a process in both markets and firms. According to Hayek (1945), knowledge is dispersed, which suggests that the variety of views needed for exploration, on what might be possible to do, largely lies dispersed outside firms, and competition is a ‘discovery process’. If sense-making and the construction of meaning in a firm are indeed more or less firm-specific, according to its specific cognitive focus, then a problem of organizational myopia looms, with the need for organizations to employ complementary cognition from outside firms with a different cognitive focus, to profit from ‘external economy of cognitive scope’ (Nooteboom 1992). Here, inter-firm collaboration begins to form an integral part of the theory of the firm. This issue yields the focus of the present paper, with the following questions:

1. What constitutes the cognitive identity of a firm
2. What is the connection between intra- and inter-firm sources of innovation
3. What are the implications for boundaries of the firm and limits to firm size

Here, the firm arises not primarily for reasons of contracting, as in contractual theories of the firm (Alchian and Demsetz 1972; Williamson 1975, 1985; Hart 1995), although in my account motives for action still form a key issue, but, more fundamentally, for cognitive reasons. For utilization and combination of knowledge, there must be some constraint on the cognitive scope of the firm. One needs focus to see. My argument comes close to that of Richardson (1972, 2002), who argued that similar, closely complementary activities are best combined within a firm, while dissimilar ones, based on different perspectives and knowledge bases, are best undertaken in different
firms. It will argue this leads to limits of firm size, in contrast with Penrose, who denied such limits.

The paper proceeds as follows. In a first section, it summarizes the views of Penrose, some points of criticism, and resulting issues for an extension of Penrosian theory. Second, it summarizes the theory of cognition used, and the derived notion of ‘cognitive distance’. Third, it sets out the notion of organizational cognitive focus, to establish ‘optimal cognitive distance’. Fourth, it derives implications for the role of inter-organizational relations for innovation, the boundary of the firm, and limits to firm size, in a trade-off between integrating capabilities in a firm and connecting capabilities between firms.

Penrose: limits and extensions

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 not objective reflections of reality, vary between people, and are therefore idiosyncratic to some extent. The present article similarly adopts a cognitive constructivist view, yielding variety of cognition between people.

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 characterized by change and uncertainty’ (Foss 2002: 156). Also, that literature did not implement Penrose’s view of cognitive differentiation between people.

However, Penrose’s account of the growth of the firm is primarily one of managerial learning, in the discovery and utilization of as yet unutilized potential of existing resources. Penrose did suggest that this moves on to the development or adoption of new resources. However, the analysis does not proceed very far in showing how that is done. In other words, ‘dynamic capabilities’ (Teece and Pisano 1994; Teece, Pisano and Shuen 2000) are assumed rather than analysed. To use the terminology of exploitation and exploration (March 1991), Penrose showed how firms learn new ways to exploit resources, but hardly showed how exploration of new resources takes place, and what the problems and limits of that are, within the firm. Goshal et. al. (2002: 291-292) distinguished between Penrosian growth, in what the firm can do, and Schumpeterian growth in what would be possible to do. So, one question for research is how firms
proceed from Penrosian to Schumpeterian growth. So far, the literature has offered limited insight in how that is done. The present paper will connect with this issue only in part, and will not give a full account of dynamic capabilities.

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 2002a). 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 conglomeration 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, from the perspective of organization as a cognitive focusing device, and from a perspective of dynamic capability, that 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 of firms.
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 (Lazonick 2002). For Penrose ‘.. organizational learning means managerial learning’ (Lazonick 2002: 270). 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.

Penrose’s pre-occupation with the role of managers also entails a ‘.. more general absence of “labour” … indeed, (the) absence of any conflicting interests’ (Pitelis 2002b: 312). While contractual theories of the firm are obsessed with problems of incentive alignment, to the neglect of learning, Penrose focuses on learning, to the neglect of problems of governance. However, a theory of the firm should cover both issues of competence (understanding, knowledge, capabilities, learning) and issues of governance (incentive alignment, conflict prevention or resolution). The distinction between competence and governance is similar, if not identical, to Searle’s (1995) distinction between the regulative order (governance) and the constitutive order (competence). In my theory both are included in organizational focus, which has a competence side, for mutual understanding, and a governance side, for ability and willingness to collaborate.

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.

**Embodied cognition and cognitive distance**

Penrose’s view of cognition as constructed and differentiated between people, and its implications for the role of management in creating ‘images’, inspired, among others, by Boulding (1956), was not elaborated in her book. Such an elaboration is offered here, in a constructivist, ‘embodied’ view of cognition. According to embodied cognition, cognition is rooted in brain and body, which are in turn embedded in their external
environment. In the present context we have two levels of embedding: people in organizations, and organizations in markets and institutions.

A possible misunderstanding of terminology should be eliminated from the start. Here, the terms ‘knowledge’ and ‘cognition’ have a wide meaning, going beyond rational calculation. They denote a broad range of mental activity, including proprioception, perception, sense making, categorization, inference, value judgments, and emotions. Following others, I see cognition as ‘embodied’, by which cognition and emotion (such as fear, suspicion), and body and mind, are closely linked (Merleau-Ponty 1942, 1964; Damasio 1995, 2003; Nussbaum 2001).

The idea that knowledge is mentally constructed from experience, yielding different results for different people, goes back to the developmental psychologists Vygotsky (1962) and Piaget (1970, 1974), with their idea that ‘intelligence is internalized action’. In the literature on management and organization, this is known as the ‘activity theory’ of knowledge (Blackler 1995), and was inspired also by the work of Kolb (1984). While the views of Piaget and Vygotsky were based on experiments in cognitive development, mostly with children, the constructivist, embodied cognition view has more recently been underpinned by Edelman’s (1987, 1992) work in the adaptive, selectionist construction of neural nets, called ‘neural Darwinism’. See also Lakoff & Johnson (1999).

In the theory of embodied cognition it has been recognized (see especially Damasio 2003, inspired by the work of Spinoza), that cognition is largely driven by feelings, which makes cognition diverse and volatile, and often limitedly coherent, without a clearly identifiable, stable, mental identity of the ego, and that such identity, in so far as indeed present, is due, in large part, to the body as a coherent source of feelings and their underlying physiology.

Briefly, the embodied cognition view claims that people perceive, interpret and evaluate the world according to mental categories (or frames or mental models) that they have developed in interaction with their social and physical environment. Since the construction of cognition takes place on the basis of interaction with the physical and social environment, which varies between people, ‘different minds think different things’, as was recognized by Austrian economists (Lachmann 1978), and there is ‘cognitive distance’ between people to the extent that they have developed their cognition in different environments (Nooteboom 1992, 1999). This connects with Hayek’s view of localized, distributed knowledge. As a result of context-dependent cognitive structuring, cognition is 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.

1 I am aware of the criticism of Piaget’s views and methodology of research (cf. Flavell 1967). However, I still think that some of his basic intuitions and ideas are valid. Apart from methodological criticism of Piaget’s work, a substantive point of criticism is that Piaget’s view is under-socialized. Here, there was an interesting difference of interpretation between Piaget and Vygotsky. In language acquisition by children, a phenomenon on which Piaget and Vygotsky agreed was that at some point children engage in egocentric speech, oriented towards the self rather than social others, and that this subsequently declines. Piaget interpreted this as an outward movement from the self to the social other; a ‘decentration’ from the self. Vygotsky ascribed it to a continued movement into the self, in an ongoing process of formation and identification of the self and development of independent thought. The reason that egocentric speech declines is that overt speech is partly replaced by ‘inner speech’. I think Vygotsky’s interpretation is the correct one.
Organizational cognitive focus

The core of my cognitive theory of the firm is as follows. Cognitive distance between people, resulting from variety of experience, presents both a problem and an opportunity. The opportunity is that variety of cognition is a source of innovation. The problem is that to the extent that cognition differs, it is more difficult to understand each other and to collaborate and utilize opportunities from cognitive variety. Note that, cognition being a wide concept in this paper, cognitive distance entails both difference in intellectual knowledge and difference in feeling and morality. Cognitive distance yields not only a difficulty of mutual understanding, or limit to absorptive capacity (Cohen and Levinthal 1990), but a wider difficulty of collaboration, including a mismatch of moral and motivational aspects of collaboration. In other words: distance includes issues of both competence and governance.

Optimal collaboration requires a trade-off between the upside and the downside of cognitive distance, seeking an ‘optimal cognitive distance’, large enough to offer variety for innovation, and small enough to enable collaboration. Here, I propose, lies the fundamental purpose of firms, and organizations more in general. That purpose, I propose, is achieved by means of an organizational cognitive focus, which has both intellectual and moral/emotional features. To achieve such focus, organizations develop their own specialized semiotic systems, in language, symbols, metaphors, myths, and rituals. This is what we call organizational culture. This differs between organizations to the extent that they have different goals and have accumulated different experiences, in different industries, technologies and markets. 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 national or regional culture). Thus the market has the higher potentiality of variety of performance, and the firm has the higher actuality of performance.

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.

Organizational focus also has a function of both selection and adaptation. In selection, it selects people, in recruitment but often on the basis of self-selection of personnel joining the organization because they feel affinity with it, and adaptation, in the socialization into the firm, and training, of incoming personnel. To perform these functions, focus must be embodied in some visible form. Such form is needed for several reasons. One is to function as a signaling device to outsiders. That is needed as a basis of the (self)selection process of incoming staff, and for recognition and identification by other stakeholders, such as customers and suppliers. Another, deeper reason is to stabilize the mental processes associated with organizational focus. As such, materialization of organizational focus has the same function as the body has for individual cognitive identity. Cognitive activities in an organization require some embodiment to crystallize, direct and stabilize cognition and communication within the organization. Here we find
symbols, such as logo’s, and style of advertisement and external communication. More for the internal function of coordination, we find the exemplary behaviour of organizational heroes, often a founder of the organization, corresponding myths, and rituals. More formalized forms of organization are procedures, for reporting, decision-making, recruitment, contracting, and the like.

An important more formal organizational form is legal identity, aimed at securing the interests of different stakeholders. Legal identity varies with the focal stakeholders and their interests. Legal identity is needed to regulate ownership and decision rights, liability, contracting, and the like. Here, firms distinguish themselves from organizations more generally. A firm is defined as an organization of capital and labour aimed at profit, in contrast with, for example, a foundation that is not aimed at profit. The legal identity of firms varies according to the regulation of liability, ownership, availability of shares, employment status, tax, and the like.

Let me contrast the present cognitive theory of the firm with earlier, contractual theories (Alchian and Demsetz 1972; Williamson 1975, 1985; Hart 1995). The latter look at organizations as systems of contracts or material incentives, to control opportunism. However, increasingly it is has been recognized that for a variety of reasons ex-ante incentive design is problematic. Due to uncertainty concerning contingencies of collaboration, and limited opportunities for monitoring, ex ante measures of governance are seldom complete, and need to be supplemented with ex-post adaptation. Such uncertainties proliferate under present conditions of professional work and rapid innovation. Professional work is hard to monitor and evaluate, and requires considerable autonomy for its execution. Rapid innovation increases uncertainty of contingencies and makes formal governance, especially governance by contract, difficult to specify. If such specification is nevertheless undertaken, it threatens to form a straightjacket that constrains the scope for innovation (Nooteboom 1999). Furthermore, the attempt to use contracts to constrain opportunism tends to evoke mistrust that is retaliated by mistrust, while in view of uncertainty there is a need to operate on trust more than on contract (Nooteboom 2002). Organizational focus, provided by organizational culture, yields an epistemological and normative ‘background’ for ex-ante selection of staff to suit organizational focus, and for ex-post adaptation, as a basis for coordination, mutual understanding, mutual adaptation, decision-making, and conflict resolution.

The content and boundaries of focus

My basic claim concerning the boundary of an organization is that it is determined by organizational cognitive focus. Later I will argue that this entails limits to firm size, contrary to Penrose’s theory. Here, a key question, considered in this section, is how far organizational focus goes, or should go. This paper is too short to give an account of all that could be said about organizational focus. For example, I do not have the space to discuss the origins, development and stability of focus. However, I will discuss features of focus that relate to its limits and boundaries, in preparation of a later discussion of the boundaries of the firm.

If organizational focus serves to seek optimal cognitive distance, what is that optimal distance? That depends on whether a firm is oriented more towards exploitation or
towards exploration (March 1991, Nooteboom 2000). In a highly stylized way, this is illustrated in Figure 1.

If ability to collaborate declines with cognitive distance, say linearly, and novelty value increases with it, say linearly, and performance is proportional to the mathematical product of the two (potential x ability to utilize it), then performance is an inverted-U shaped function of distance, yielding some optimal distance (Nooteboom 1999). Now for exploitation (Figure 1b), which is oriented towards efficiency, in a fine tuning of complementary capabilities, where lack of error or mismatch is more important than novelty, the marginal utility of novelty is less (lower positive slope of the novelty line) than for exploration (Figure 1a), which is oriented at more radical novel combinations, and the marginal disutility of lack of understanding and ability to collaborate is greater (higher negative slope of the ability line). As a result, as illustrated in Figure 1, optimal cognitive distance is lower for exploitation than for exploration. In exploration cognitive distance has more relative advantage. This also illustrates the problem of combining exploitation and exploration in a single organization: there is a tension between the needs for small and for large cognitive distance at the same time. One way for firms to solve this problem is to specialize in either of the two, and engage in an alliance with a firm that specializes in the other. A classic example of this is small biotechnology firms that focus on the exploration of novel substances or processes and collaborate with large pharmaceutical companies for their exploitation.

Now, concerning the boundaries of the firm, and limits to firm size, the point will be that even for exploration there must be some limit to cognitive distance, for the sake of exploiting its potential. After this highly stylized representation, let me unpack the notion of organizational focus and its limits.

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 organizations varies. 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. For exploitation focus needs to be tighter, and for exploration more loose.

Fourth, focus may have different content. In particular, on the governance side it may entail 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. 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. The first occurs more in exploitation and the second more in exploration.

Fifth, and this will turn out to be a central point, 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. 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 2. Here, for simplicity of exposition, the intermediate level of culture 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 flexibility and speed of action, but it reduces variety of cognition on the deep level. It entails more indoctrination. Thus efficient exploitation is enhanced by deep level coordination, and exploration is constrained by it.
I will argue that organizations serve especially to coordinate on the deep level, with an advantage of easier and faster understanding and agreement, to enable exploitation, while collaboration between organizations operates more on the surface level, with the advantage of greater variety on the deep level, allowing for a wider scope of exploration. Organizational focus entails a certain myopia, which can be compensated with external relationships between firms, at greater cognitive distance. Here, the theory of the firm includes a theory of inter-firm relationships.

The notion of cognitive distance 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 entails 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.

**Boundaries of the firm**

The point of moving from specific activities to more generic capabilities, proposed by Penrose, 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, require renewed investment in mutual understanding and ability to collaborate. 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, beyond linkages that are currently operational. This is again an argument of speed and flexibility. This 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 is further sharpened with an analysis of the depth of cognitive focus. For the same reason that a focus on capabilities yields more speed and flexibility of new activities than a focus on activities, cognitive alignment in deep structure rather than in surface regulation is more generative and flexible, yielding more flexibility and speed in re-configuring capabilities. This was illustrated in Figure 2. Now, deep level categories are more tacit than surface level rules, and hence require more time to develop and investment in them is more specific, requiring longer time to recoup, 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. Note that here some of the logic of transaction cost theory, in terms of specific investments, creeps in.
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 and deeper level 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 boundaries of the firm and limits to firm size. 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.

**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, between different levels of organization. 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 range of capabilities, in each of them actual variety decreases due to rising needs of common understanding and agreement. 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. That is better done, with more flexibility and efficiency, in ad hoc collaboration between firms, than within a firm.

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.

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 smooths 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. In Figure 1, an upward shift of the line for ability to collaborate yields an increase of optimal cognitive distance, as illustrated in Figure 3.

Postrel (2002) asked when communities of practice should invest in shared knowledge (reduce cognitive distance), 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 (i.e. absorptive capacity, ability to collaborate), then they can thereby achieve fit to whatever other communities do.

The principle is illustrated in Figure 4, which is derived from Figure 2. 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 again 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, an ongoing 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. One would want to leverage the investment in a wider range of potential collaborative relationships. 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, first of all perhaps, that it is not easy to refute Penrose. For every possible 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.

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 that extends 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 firms 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 of mutual understanding (absorptive capacity) and ability to collaborate, and of 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

Alchian, Armen A. and Harold Demsetz

Berger, Peter L. and Thomas Luckmann
1966 The social construction of reality: A treatise in the sociology of knowledge, Garden City NY: Doubleday.

Blackler, Frank

Boulding, Kenneth E.

Cantwell, John

Choo, C. W.

Cohen, Michael D. and Dan A. Levinthal

Damasio, Antonio. R.

Damasio, Antonio. R.

DiMaggio, Paul J.

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

Edelman, Gerald
Edelman, Gerald

Flavell, John H.

Foss, Nicolai J.

Fransman, Martin

Ghoshal, Sumantra, Martin Hahn and Peter Moran

Hart, Oliver. D.

Hayek, Friedrich A. von

Kay, Neil

Kogut, Bruce and Udo Zander

Kolb, David

Lachmann, Ludwig M.

Lakoff, George and Mark Johnson
Lazonick, William

Loasby, Brian

Loasby, Brian

March, James G.

Merleau-Ponty, Maurice

Merleau-Ponty, Maurice

Nelson Richard R. and Sidney Winter

Nooteboom, Bart

Nooteboom, Bart

Nooteboom, Bart

Nooteboom, Bart

Nussbaum, Martha C.
Patel, Pari and Keith Pavitt

Patel, Pari and Keith Pavitt

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

Piaget, Jean

Piaget, Jean

Pitelis, Christos (Ed.)

Pitelis, Christos

Porter, Michael E.

Porter, Miachel E.

Postrel, Steven

Quinn, James B.

Richardson, G.B.
Richardson, G.B.

Richardson, G.B.
2002 Mrs. Penrose and neoclassical theory, in C. Pitelis (Ed), The growth of the firm; The legacy of Edith Penrose, Oxford: Oxford University Press: 37-44.

Ring, Peter S. and Andrew van de Ven

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

Scott, W. Richard

Searle, John R.

Schumpeter, Joseph A.
1934 Capitalism, Socialism and Democracy, London: Unwin.

Simmel, Georg

Simon, Herbert A.

Smircich, Linda

Teece, David J. and Gary Pisano

Teece, David J., Gary Pisano and Amy Shuen
Vygotsky, Lev

Weick, Karl F.

Weick, Karl F. and Karlene H. Roberts

Williamson, Oliver E.

Williamson, Oliver E.
1985 *The economic institutions of capitalism; Firms, markets, relational contracting*, New York: The Free Press.

Witt, Ulrich

Zollo, Maurizio and Sidney G. Winter
Figure 1  Exploitation and exploration

Figure 1a  Exploration

<table>
<thead>
<tr>
<th>ability to collaborate</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>performance</td>
<td></td>
</tr>
</tbody>
</table>

optimal cognitive distance

Figure 1b  Exploitation

<table>
<thead>
<tr>
<th>ability to collaborate</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
</tr>
<tr>
<td>performance</td>
</tr>
</tbody>
</table>

optimal cognitive distance

Figure 2  Levels of coordination

<table>
<thead>
<tr>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface level regulation</td>
</tr>
<tr>
<td>Deep level cognitive structure</td>
</tr>
</tbody>
</table>
Figure 3  Upward shift of ability to collaborate

Figure 4  Absorptive and collaborative capacity

- A and B: narrow collaborative capacity and small cognitive distance
- A and B: wide collaborative capacity and large cognitive distance