

ABSTRACT. Associated with effects of scale, scope, experience and learning there are effects of firm size on transaction costs; in the stages of contact, contract and control. These effects are due to "threshold costs" in setting up contacts, contracts and governance schemes, and to differences with respect to the factors that cause transaction costs: bounded rationality, opportunism, uncertainty and transaction specificity of assets. Implications are considered for firm strategy and public policy.

1. Introduction

While small firms often have behavioural advantages, in many activities they are at a disadvantage with respect to costs. In costs, there are effects of scale, scope, experience and learning. Small firms generally produce small volumes (scale) of few products (scope). Often, they have not been in business long and thereby have little benefit from economies of experience. Often they have limited capacity for the acquisition of knowledge (learning).

Effects of scale, scope, experience and learning play a role in the decision whether to produce an input that one requires, or to purchase it from an outside producer ("make-or-buy decision"). Transaction Cost Economics (TCE), initiated by Coase (1937) and developed further by Williamson (1975, 1985), is designed as an answer to the question why there are firms and why different activities are combined in one firm. There are firms, rather than only individuals selling the fruits of their individual labour, because of economies of scale in production. Firms may produce not only one but several products, including inputs re-

quired for a given product, because of inseparable economies of scope between the different products. This yields an argument in favour of inside production. On the other hand, according to Williamson economy of scale favours outsourcing because a specialized outside producer, serving multiple users, can produce at a higher volume and hence more cheaply.¹ However, there are also costs of transaction, and these may exceed the benefits of outsourcing, thus favouring the decision to make. Effects of experience yield an obstacle in the shift from an established to a new producer, and hence may provide an obstacle to a shift either way: from inside to outside production or vice versa. Problems of learning, or the development of new competence, are relevant because they yield an added value to inter-firm cooperation.

In its consideration of the make-or-buy decision, standard TCE considers the effects of scale and scope in costs of production, and to a limited degree their effects in costs of organization, but hardly their effects in costs of transaction. TCE tends to ignore effects of experience and learning. The purpose of the present article is to fill these gaps: what are the effects of scale, scope, experience and learning, and associated with these the effects of firm size, on transaction costs, and what are the implications for make-or-buy decisions and other aspects of policy?

As a preliminary, a summary of TCE is supplied, a survey of types and sources of effects of scale, scope, experience and learning, and the definition of "small and medium sized enterprise" (SME) that is employed. The analysis is conducted from a theoretical perspective, but in the process many bits of experience and results from empirical study will be taken up. The article aims to provide a theoretical reconstruction of known phenomena.

In the analysis of effects of firm size in transaction costs, we employed TCE, but that does not suffice, and extended theory is required. TCE has

Final version accepted on 16 July 1992

*School of Management & Organization,
Groningen University,
P.O. Box 800,
9700 AV Groningen,
The Netherlands*

limitations in the context of rapid innovation, because it is not a dynamic theory: it does not incorporate a theory of change of preferences and learning (changing capacity for perception, interpretation and evaluation) as the transaction relation develops in time, and the role thereby of interaction between transaction partners.²

Effects of firm size on transaction costs are presented in several steps: generic effects that apply for small firms in the role of both buyer and supplier, and effects that are specific to either role. We will demonstrate that smaller firms, as both suppliers and buyers, incur higher transaction costs directly, and cause higher transaction costs for transaction partners. In the latter case, the partner may demand compensation (on the threat of going elsewhere), whereby the higher costs are reflected back to the smaller firm.

We conclude that there are pervasive effects of firm size, and we consider some of the implications for policies of firms and governments, designed to reduce or eliminate disadvantages in transaction costs.

2. Transaction cost economics

According to Williamson's prior work (1975), the user of some productive input has a stark choice to make between two "governance structures": integration of production of the input in the "hierarchy" of his own firm, or purchase in "the market", from an outside producer. The advantage of outside purchase ("outsourcing") lies in the "high powered incentives" of the market and in economy of scale, by which the input is produced more cheaply by a specialized producer supplying multiple users. There is a disadvantage due to transaction costs; particularly costs of dependency on an outsider partner, which creates vulnerability to opportunism in so far as conditions are uncertain (during execution of the transaction). To the extent that investments are more "transaction specific", there is less scope for economy of scale in outside production, because the volume of production decreases to the volume needed by the focal user, and transaction costs of dependency increase. At some point transaction costs exceed the benefits of outside production and in-house production is to be preferred (because under

hierarchical control risks due to dependency are less).

For transactions with an "intermediate" degree of transaction specificity of assets, Williamson later (1985) made allowance for two types of governance structure "between market and hierarchy". "Bilateral governance", with the institution of different kinds of safeguards to limit transaction costs, in the case of frequent, recurring and substantial transactions, which make the investment in such measures worthwhile. Examples of such safeguards are: ownership by the buyer of transaction specific productive assets; guarantees by the buyer, in the form of a guaranteed quantity and price of purchase or a severance payment; the supply of "hostages". Guarantees given by the buyer to the supplier require counter-safeguards to protect the buyer against misuse. For more incidental or smaller transactions, which do not warrant the cost of such measures, "trilateral governance" is proposed, with the appointment of a third party as an arbitrator to regulate any conflicts that may arise.

More on a macro level, North (1990) adduces transaction costs as a reason for the development of institutional arrangements, including sets of culturally determined norms (social, moral, legal, personal), to contain transaction costs that would otherwise be too high to allow for the degree of specialization on which prosperity depends.

TCE is founded on two behavioral assumptions: bounded rationality and opportunism. It is due to the combination of these two conditions that transaction costs arise to the extent that assets are "transaction specific", whereby transaction partners become dependent on each other (are "locked in"). Bounded rationality is taken to arise from the scarcity or cost of information and limited capacity for information processing. If rationality were unbounded, all possible contingencies could be foreseen, even those arising from opportunism, and could be incorporated in a contract prior to commitment. TCE does not claim that opportunism is practised by everyone all of the time, but by some people some of the time, but since *ex ante* one does not know when and by whom, governance of contracts should be designed to take opportunism into account. If there were no opportunism, contracts could be left incomplete in the trust that unforeseen contin-

gencies would be met in a spirit of cooperation and mutual benefit. In either case transaction costs would be low. Since in fact bounded rationality and opportunism both occur there will be transaction costs.

In traditional TCE, the causes of the existence of firms are sought in considerations of efficiency. Given a certain state of knowledge and technology and a preference ordering, efficiency is maximized by trading off production costs, transaction costs and costs of organization. The approach can be characterized as one of comparative statics, with a focus on static rather than dynamic efficiency. We define dynamic efficiency as efficiency in innovation, which is characterized by shifts of knowledge, technology and preference. Shifts of knowledge imply shifts of preference, since they may yield shifts of perception of outcomes of opportunities or perception of new opportunities. Preferences may also shift due to a shift of evaluations of outcomes of opportunities or of corresponding risks. A dynamic theory of transaction costs requires a theory of learning (including preference formation) which includes the role of interaction between transaction partners.³

A transaction can be seen as an event during a process of exchange which consists of three stages: contact, contract and control. In the stage of contact, the buyer incurs search costs and the seller incurs costs of marketing. In the stage of contract, costs are incurred in the preparation of an agreement to transact, in which one tries to anticipate possible problems during execution. These costs include costs of search of information on: the reliability of the transaction partner, possible contingencies in the future, degree to which investments will be sunk. They further include costs of negotiation, legal advice, set-up of arbitration, design of safeguards (and guarantees against their misuse), etc. In the stage of control, after the commitment to transact has been made, there are costs of: monitoring, settling disputes ("haggling"), renegotiation, arbitration, litigation, loss of investments due to the relation breaking up, etc.

3. Scale, scope, experience and learning

To take into account effects of firm size, we need to consider effects of scale, scope, experience and

learning. The diversity of these effects tends to be underestimated, and therefore we provide a brief survey of their different types and sources.

3.1. Effects of scale

Economy of scale is defined as the decline of average costs (per unit of product) with an increase of production volume per unit of time, where production capacity is variable. Note that the effect depends on the rate of production (a flow variable).

Effects of scale occur in different areas: production, staff support, management, R&D, finance, distribution, marketing. They occur on different levels of aggregation: production unit, establishment or plant, enterprise. While in several areas of production effects of scale have lessened, allowing for smaller scale production units or plants, in many issues of finance, R&D and marketing (distribution, brand names) effects of scale have increased.

Sources of economy of scale are: indivisibilities of people and facilities, specialization and laws of mathematics and physics. Indivisibility yields so-called "threshold costs": no matter how small output may be, there is a minimum capacity of people or facilities.⁴ Examples are: an attendant at a counter or check-out in a shop, an operator at a console in a chemical factory, the driver/pilot of a truck/plane, the supervisor of a team, a teacher before a class. An example of a minimal capacity of a machine is a truck (room for the driver, motor, and at least four wheels), a shop (room for a check-out or counter plus some minimal space for queuing, depending on the climate). Important reductions of effects of scale in production are due to technical developments that reduce minimum size of machines; in computers and electric motors, for example. As a result, threshold costs of labour have become relatively more prominent, yielding an incentive to reduce or eliminate them by automation of service activities (dispensing of money and goods, observation and monitoring, answering telephones). But there are also minimum set-up costs of machines, contracts, the granting of a subsidy, advisory services. It is particularly this latter source of economy of scale that is relevant to transaction costs. An example of a law of mathematics as a source of economy of

scale is the well-known formula relating the surface of a sphere (or cube) to its content: surface is proportional to content to the power $2/3$. This is relevant in several industries: chemical industry ("pots and pans"), housing, transport facilities. There, capacity is related to content, and cost to surface: material costs and hence weight and cost of transport, cost of cleaning, cost of heating. Thereby, costs increase less than proportionately with capacity.

3.2. *Effects of scope*

Effect of scope is defined as a reduction of average costs (per unit of product) by the addition of another product to the portfolio of production. Effects of scope also arise in different activities and on different levels: production, management, R&D, staff support, finance, marketing. Sources of economy of scope are: indivisibilities of resources, various forms of complementarity, and interaction in production. The sharing between different products or activities of indivisible facilities (buildings, support services, distribution channels) yields a scope which is not "inseparable": different firms may share the facilities or may utilize them only to the extent required, by renting capacity only for the time needed.

Complementarities are quite varied. There is complementarity in materials: the separation of hide (leather production) and carcass (meat production); the separation of tar (for roads), naphta (for the chemical industry) and different grades of gasoline from crude oil. There is complementarity in risks: production for export and import markets; for industrial and consumer markets; for markets that benefit from wage increases and markets that benefit from lower wages. There is complementarity in time: production in markets that mutually compensate peaks and troughs in demand (summer and winter, week and weekend, day and night). There is complementarity in brand names: an additional product benefits from a brand name established for existing products (but also increases vulnerability: a miss of quality in one product contaminates other products under the same brand name). Some of these complementarities are more inseparable than others. Brand name complementarities are separable at the plant level but not on the enterprise level. For share-

holders, risks can be diversified outside the firm, in a stockmarket portfolio. The owner/manager, however, often has all his stakes in his own firm.

"Interaction" requires coincidence in time and place and yields strong inseparability. Examples are the interaction between labour (operator) and capital (machine), and interaction within labour (teamwork). Examples of inseparability of scope in teamwork are: loading furniture onto a removal van, an operating team, a symphony orchestra.

3.3. *Effects of experience*

Effect of experience is defined as the decline of average cost (per unit of product) with an increase of cumulative uninterrupted production. Its difference from economy of scale is that the latter depends on current production volume per unit of time (flow variable), while experience depends on accumulated production since the start (stock variable). It may also arise in different activities and on different levels: in production, support services, marketing, finance.

In production (and support) the source of the effect is that one becomes increasingly expert, by a cumulation of knowledge and a honing down of processes to essentials, discarding redundant activities. In finance, it entails reputation of continuity, generating lower perceived risk and thereby cheaper and easier credit. The experience effect carries a pitfall: as the advantage of experience increases, it becomes more difficult to surrender this advantage in a shift to new technology. Thus experience may lead to conservatism.

The experience effect is the outcome of doing more of the same, whereby one eliminates redundant activities. It is a process of elimination. Learning to conduct new activities (inefficiently at first, due to lack of experience), although contingent upon experience, is a different process. It is a process of expansion.

3.4. *Effects of learning*

Effects of learning, in the sense of learning new things, have been neglected in the relevant literature. It would take too much space to fully develop the theory here (see Nooteboom, 1992). Very briefly: perception, interpretation and evaluation are performed on the basis of categories that

condition knowledge in the double sense of enabling and limiting it ("you can't look in all directions at the same time"). Such categories are developed in interaction with one's (natural and social) environment, and are thereby context- and pathdependent. Associated with this, knowledge is partly "tacit": we perceive, interpret and evaluate according to categories that we are not even aware of (cf. Polanyi, 1962). This yields a difficulty for self-criticism, even if one is dedicated to it. As a result one runs the risk of failing to perceive, understand and appreciate developments that present threats or opportunities. This is more important to the extent that the environment of firms is turbulent, as is presently the case. From this perspective there is an added value in transaction relations, in that from a different but relevant perspective and experience suppliers or customers may serve to signal developments or provide new competences that would otherwise be missed. The analysis is consistent with Granovetter's (1982) notion of the "strength of weak ties", i.e. of ties with partners who have backgrounds and perspectives that are different from one's own.

4. Firm size

In studies of SME, a vexing question always is what definition one should employ. For statistical purposes, the boundary between small and medium-sized business is variously drawn, in different countries or different contexts within a country, at 10, 20 or 50 persons engaged, at the enterprise level, and the boundary between medium size and large enterprise at 100, 200 or 500 persons engaged. Other boundaries can also be found. Often, the boundaries are drawn differently for manufacturing and services, because a shop is sooner considered "big" than a factory. The universally "best" demarcation, regardless of context, is an ephemeral concept. The boundary one would like to draw depends on the type and purpose of the study. Universal demarcations serve the same purpose of a convention as traffic rules or rules of accounting: their only virtue, but it is not a small one, is that they are universal, i.e. that they are adhered to by everyone (at least within a country). The virtue of a conventional demarcation of SME is not that it claims to be the "true" one, but that it does make more or less

sense and by adhering to it at we are at least all talking about the same thing. Here, we adhere to the Dutch statistical convention: small firms less than 10 and medium-sized firms in between 10 and 100 persons engaged. That will in many if not most cases mean that small firms, and in manufacturing industry also medium-sized firms, are single-establishment firms.

It is always a delicate matter to make general statements about SME, because its diversity may be its most important characteristic. This yields a dilemma for public policy. On the one hand measures should be differentiated, meeting diversity, in order to be effective. On the other hand such differentiation raises the level of complexity of institutions and conditions that is already a problem for SME in view of the limited cognitive, information processing capacity of many small firms. In view of this dilemma, it is of some importance to keep on looking for general statements about SME that are warranted in spite of its diversity.

Compared with large firms, small firms have potential advantages and disadvantages. Generally, their advantages are behavioural (in human resources) and their disadvantages lie in material resources (cf. Rothwell, 1985, p. 9). Potential behavioural advantages are: entrepreneurial drive and risk taking, motivation and perseverance, motivated labour due to lack of bureaucracy and specialization, flexibility, proximity of management to customers, proximity of management to the shop floor. Whether these potential advantages are in fact realized depends on the motivation and orientation of the entrepreneur, which vary widely. The firm bears the personal stamp of the entrepreneur in many respects: on the primary goal of the firm (profit, growth, stability, job satisfaction); on orientation (technical, commercial, social); on style of internal and external communication, work conditions, etc. Some small firm owners are motivated towards "true" Schumpeterian entrepreneurship, others towards maintaining established, traditional ways of doing things (craftmanship), maintaining independence, staying small, having a quiet life. They can afford to entertain unorthodox objectives (unorthodox in the view of economic science) to the extent that they own the firm, which they often do.

There are several disadvantages in resources,

including high costs due to diseconomies of small scale, limited scope, experience and learning. In small firms the entrepreneur often participates intensively in day to day production, on the shop floor. As a result, his time is extremely scarce. In medium sized firms operational tasks have been fully delegated, and management tasks partly. Small firms will in majority have no specialized staff for finance, personnel or marketing, and certainly not for legal affairs, while in medium sized firms such support will be present only to a limited extent. As regards level of training/education of the entrepreneur and his staff, the conventional wisdom, based on empirical research, is that the level is lower than in large business. For the Netherlands, detailed statistical research by Bol, Nijssen, Regter and van Rossum (1982) shows that smaller business is less represented at both the top of the scale (higher education) and the bottom (unskilled labour). Increasingly, it is important to make a distinction between technically more advanced firms, particularly in new areas of technology, with high levels of education, and more traditional firms in more traditional industries, with a lower level. The former are often spin-offs from large firms, or are founded fairly recently by graduates from universities or polytechnics.

Small firm size correlates strongly with small scale and scope, and less strongly with limited experience and limited learning capacity. In none of these respects there is one-to-one correspondence. A firm which is small in terms of enterprise may be not so small in terms of establishment or plant. A retail enterprise of 50 people may be called medium-sized, or even small, depending on the statistical convention that one chooses to adopt, but in many trades, a shop of 50 people is quite large. If an establishment is small, a technology may be available in which scale effects are small. However, by definition a small firm incurs effects of scale on the enterprise level. Small firms may compensate for lack of economy of scope by cooperation with other firms, but this requires forms of quasi integration to the extent that the economy of scope is only partly separable. A well known form of such quasi integration is the arrangement of "industrial districts". A small firm may have a long experience, yielding a moderate or even high cumulative production in spite of a modest volume of production per unit of time, due

to low or zero growth. But in such small, experienced, long established firms there is a threat of technological conservatism. But many small firms will be relatively young. In SME, educational level and capacity to detect and absorb relevant new information (learning) can be relatively high even compared with large firms, but this is an exception rather than a rule.

Often small firms manage to achieve success by exploiting their strengths and compensating for their weaknesses. Exploiting strengths in application-oriented innovations, market niches and flexibility. Compensating for their weaknesses by means of networks of external relations or related structures (industrial districts).

Summing up: small firms often lack economies of scale, scope, experience and learning, though not universally in all those aspects, and in varying degrees. They may compensate for such disadvantages in several ways, notably by exploiting their behavioural advantages, but this does not deny that those disadvantages occur.

In the present context of transaction costs we first consider generic effects of firm size, that apply to both suppliers and buyers, due to threshold costs of transactions and size effects in the issues of bounded rationality, opportunism and uncertainty. Since the transaction specificity of assets is related to the roles of buyer and seller, we will consider that dimension later.

5. Threshold costs

There are threshold costs in each of the stages of contact, contract and control: to set up and execute an appointment with a potential transaction partner, to judge its perspective, to make or judge an offer, to set up a contract, to set up channels of communication, and to set up a scheme for control, whether in the form of guarantees and controls in bilateral governance or in the form of an arbitrator in trilateral governance. These threshold costs arise regardless of the size of the transaction, and thus weigh more heavily for the smaller transactions associated with smaller firms, for both suppliers and buyers.

6. Bounded rationality

Bounded rationality is important in TCE because

in the stage of contact it limits opportunities of search, in the stage of contract it yields imperfect contracts, to the extent that future contingencies are uncertain, and in the stage of control it limits the capacity to monitor performance. A larger firm, with more markets for a given product and/or more products, tends to face a wider set of relevant contingencies, and to command a wider set of available options for action. The wider sets of contingencies and actions of the larger firm require a greater capacity for the collection and processing of information. The wider set of available actions imposes a demand for greater information processing capacity on its transaction partners, because they need to explore the threats of opportunism which they might have to meet.

The demand for greater capacity of information processing in the larger firm is met by a greater ability to identify, collect and absorb relevant external information, because of the possibility to employ specialized staff in different functional areas: marketing, finance, personnel, legal matters, technology, R&D, planning, etc. Here, small firms are at a disadvantage. Also, often though certainly not always, the capacity for absorption of external information is limited due to a relatively low level of education or training of the entrepreneur and his staff. In the class of small firms in more advanced, science dominated sectors this latter point does not apply. An additional disadvantage for the smaller firm is that the perspective from which external scanning is performed is often dominated, and thereby restricted, by the personal perspective of the entrepreneur. This is particularly important because the entrepreneur often has a bent towards stubbornness, sometimes to the point of obstinacy and blindness. Note however, that stubbornness is close to perseverance, which is a virtue in the struggle to enter a market and survive. Here, limited resources of rationality may go together with a behavioural advantage.

Summing up: in the smaller firm rationality is more bounded along three dimensions: width (fewer functional areas in staff support), depth (lower level of education, with the exception of firms in science based sectors) and variety (dominance of the personal perspective of the entrepreneur).

The use of external suppliers of business services (on strategy, technology, export, market-

ing, legal issues, financial planning and control) can compensate for the lack of internal specialist staff. This is consistent with the empirical evidence that small business managers spend relatively much time in developing and maintaining external sources of information (see, for example, Johnson and Kuehn, 1987; Smeltzer and Fann, 1989. For the Netherlands: NOBIN, 1984). But note that this entails second order costs of contact (search), contract and control. In all stages of the second order transaction, costs are high since the service product concerned is typically a "credence product": its quality is difficult to judge even after consumption. If one were able to judge quality well, one would possess the expertise whose lack is precisely the reason for seeking it outside. Here, trust is involved in a dual sense: in disinterestedness and in competence. Disinterestedness entails that the provider of information has no axe to grind and is objective and impartial. Competence entails that he commands the necessary technical expertise and the experience in applying it to a smaller firm, with an understanding of its problems and priorities. The latter aspect of competence is often lacking in sources such as government or semi-government agencies and universities. To judge all this, the entrepreneur may have to employ yet other outside contacts, to delegate judgement or to gather advice, for the contact, contract and control regarding a supplier of some service, and this involves him in yet a further round of third order transaction costs. For this problem, entrepreneurs often employ a personal network of long standing relations with trusted family, colleagues, accountants, customers, local politicians, suppliers or the bank. As a result, as argued by Johannisson (1986), this network may be one of the major assets of the entrepreneur, although it is intangible. The analysis explains why in much empirical research informal network contacts turn out to absorb relatively much time.⁵ Thus, to some extent small firms manage to reduce or solve the problem of bounded rationality and lack of information, but at a considerable expense of time and effort in networking.

The costs of the second order transactions with external professional advisors are further increased by threshold costs in the stages of contact and contract that were identified earlier. The set-up costs often yield a minimum, fixed charge,

whereby the service is more expensive for the smaller firm, or in the absence of a minimum charge the supplier is reluctant to supply. Examples abound: a minimum charge for insurance, a government subsidy only for firms above a certain size, refusal to supply below a certain volume, price reductions for higher volumes, etc. To reduce threshold costs, suppliers of business services to small firms may strive for standardization of procedures, to be applied with limited variation to groups of firms, thus yielding more ready-made than tailor-made advice, as discussed in Nooteboom, Zwart and Bijmolt (1992). This encounters complications. Firstly, for several reasons it is precisely among smaller firms that the diversity of problems and needs is greatest, requiring more not less tailor made advice. Secondly, standardization yields complications of coordination between competing suppliers. If standardization does not also take place across competing suppliers, the costs of search may further increase for the small business user. If standardization across competitors is sought, this raises familiar problems of such collective action, which we will not discuss here.⁶

A third reason for high second order transaction costs, in business services to the smaller firm, is that in small firms there is often a lack of the records that may be needed to provide the service: on orders, deliveries, services provided (installation, training, maintenance, repairs), payments received (corrected for any rebates), receivables, payments made, inputs, stocks, throughputs, labour hours, sickness and other absence, training, quality control, waste and scrap, defects and rejects, returns, customer complaints, internal maintenance and repair etc. The proper response may not be the institution of more formal records and procedures, with the purpose of reducing second order transaction costs. The records may be justifiably lacking because they are relatively expensive (due to economy of scale in their set-up costs), and because in fact their lack entails a strategic advantage. For operational purposes they may not be needed due to effective and cheap arms-length, informal oral communication, visual inspection, personal memory, etc. If due to small size the firm does not need formal procedures and records for internal coordination, their lack contributes to flexibility and speed of response, due to

lack of bureaucratic obstacles. Here, behavioural advantage may go together with limited administrative resources. The best response may be to simply accept the higher second order transaction costs.

Summing up: due to more restrictive bounds of rationality in small firms, transaction costs are higher on the level of the first order transaction, or are higher due to costs of higher order transactions, in the supply of business services and the evaluation of their quality, or due to costs of networking. Thus, one way or the other, transaction costs are higher.

We noted that particularly in conditions of turbulence there is an added value of transaction relations with a view to learning: to escape from obstacles in perception, interpretation and appreciation. This applies to both large and small firms. However, we already noted that in the smaller firm, due to personal dominance of the entrepreneur and his stubbornness obstacles to the recognition of opportunities and threats may be higher. On the other hand, in smaller firms the distance between strategic decisions and the market is shorter, which yields the opportunity of a quicker response. The net effect depends very much on the balance of characteristics in the individual firm, and therefore is largely an empirical matter. There is, however, a more systematic effect. In small firms knowledge is often more craft-like and based more on experience (learning by doing), while in large firms it is often based more on formal, more abstract learning and more explicit rules and procedures associated with the need to communicate more widely and therefore more formally. As a result in small firms knowledge tends to be more tacit, with the greater difficulty of criticism.

7. Opportunism and uncertainty

TCE considers "internal" uncertainty, concerning the conduct of transaction partners, which includes the possibility of opportunism, and "external" uncertainty, concerning the contingencies that may arise as the transaction proceeds.

As already noted, a larger firm faces a wider set of relevant contingencies, and commands a wider set of alternative actions. This yields effects of firm size in opportunism and uncertainty. The wider

sets of contingencies and options yield a spread of risks to the extent that there is no positive correlation between outcomes of different activities. This makes larger firms more robust with respect to uncertainty, and this contributes to lower transaction costs. The vulnerability of large firms to opportunism of transaction partners tends to be less because the risk can be compensated with other transactions, and this also contributes to lower transaction costs. To put the point differently: a larger firm is less sensitive to opportunistic action of a single transaction partner, to the extent that it has more transaction partners.

The wider sets of contingencies and options also offer more opportunities for the exercise of opportunism by larger firms. However, when the exercise of opportunism is visible and becomes known, reputation effects will restrain the larger firm more strongly because of the larger number of transaction relations that may be damaged by a reputation of opportunism. A firm that has built up a name as a reliable transaction partner, with limited opportunism, at the visible expense of opportunities foregone for the appropriation of short-term advantage, offers a reduction of the perceived risk of opportunism, and thereby offers lower transaction costs to current and potential partners.

Small firms cannot by themselves offer this safeguard against opportunism to the extent that they have started only recently. Also, a narrower range of transaction relations, related to a smaller volume of sales and a narrower range of products, reduces risk for the perpetrator of opportunism, and thereby increases its attractiveness and hence the suspicion of it on the part of partners. To put this point in different terms: small and in particular new firms may generate suspicions of their possibly being "fly-by-night" operators, who are in for a quick "hit and run". This contributes to a tendency that with respect to small firms perceived risk of opportunism is higher, which makes them less attractive as transaction partners or entails the need to offer some compensation. One possible compensation would consist in sunk costs, which are large relative to the size of the small firm, associated with the transaction relation, to insure the partner against the risk of a hit and run tactic. However, this increases the dependence of the small firm on his partner, raising transaction costs.

Another solution for small firms is to become member of a trade association which can yield credible safeguards against opportunistic conduct of its members. This is what we observe, but again the (partial) solution of the problem confirms rather than denies that the problem exists.⁷

A further problem for the smaller firm is that because of a lesser spread of risk, it presents a greater risk of default quite apart from opportunistic intentions: the smaller firm simply goes broke more easily, thereby yielding a greater risk of discontinuity towards its partners. Discontinuity raises transaction costs for the partner because of loss of transaction specific investments, or interruption, lower quality or higher cost of supply. This either raises the transaction costs for the partner, thus reducing the relative attractiveness of the small firm for a transaction, or the transaction costs for the small firm, in order to offer a compensation. The problem is particularly acute if politically a buyer is held ransom with respect to the employment and hence the continuity of its small suppliers. This would predict that in countries where continuity of employment is a political constraint, buyers will be less inclined to let suppliers be dependent on them for more than a certain percentage (often set at 30%) of their total sales. Here, a (partial) solution may lie in a trade association or other institution which provides safeguards against the consequences of sudden failure of its members, in the form of some fund for insurance or compensation. But this also carries costs.

Summing up: smaller firms face higher transaction costs because they are more sensitive to uncertainty, are more vulnerable to opportunism, yield a greater risk of unintended discontinuity for their partners and raise higher suspicions of opportunism. Transaction costs are higher either directly or indirectly, in the expense required to compensate for greater risks to their partners.

8. Asset specificity

After generic effects of firm size, regardless of the role of the small firm, we now consider effects of asset specificity, from the perspective of small firms as buyers and small firms as suppliers. For a proper understanding, another extension of TCE is now required.

In standard TCE the suggestion is made that transaction costs tend to be symmetric between buyer and producer, so that they are both "locked in", in case of transaction specific investments. The producer typically makes investments in assets required for production, and if those are transaction specific, he stands to lose them in case the transaction is discontinued. On the other hand, so the theory goes, in case of discontinuity the buyer also suffers because there will be delays in obtaining an input of equal quality and price. But in Nooteboom (1993) it is shown that there is carelessness here in the treatment of the concept of specificity. Implicit in the argument is the assumption that there is a one-to-one correspondence between specificity of productive assets and monopoly. Since the first entails vulnerability of the producer and the second entails vulnerability of the buyer, they are always both vulnerable, or neither is vulnerable. But although this will often be the case, it is not necessarily so. Assets can be specific to a product which is not unique: no other products can be made with the asset (and there is no alternative demand for the product), but there are other products that can serve as more or less close substitutes. This makes the producer vulnerable but not the buyer. Of course one may argue that this possibility is uninteresting because the producer would be unwise to engage upon such a transaction. But this counterargument places high demands on the producer's knowledge of the market, and ignores that there may be other characteristics of the transaction, whereby the disadvantage to the producer might be offset. Conversely, a unique product can be offered with assets that are hardly specific: one can with general purpose programmable machinery assemble unique products from standard components. One may object that then the product may easily be copied, so that it does not really yield a monopoly, but this counterargument ignores the possibility of entry barriers.

But let us consider the paradigm case, where assets are used which are specific to a unique product purchased by only one buyer. Even in that case symmetry of dependence need not obtain. The producer stands to lose an investment, while the buyer risks a temporary disruption of supply until he can acquire an alternative supplier with a substitute (or make it himself). The cost may but

need not be symmetric: this is an empirical matter, depending on the relative size and sunkness of the investment and the uniqueness of the product.

Let us turn to effects of firm size for a small scale producer. Two generic problems have already been discussed: problems due to bounded rationality and problems due to greater risk and higher perceived opportunism. In the case of the producer, costs of contact are costs of marketing: understanding market needs and opportunities, and threats from competition, new entrants, substitutes, etc. For a small scale producer these costs are relatively high. This problem can be recognized in the often reported fact that firms conduct less formal and explicit strategic planning and marketing.

A third problem for the small producer is that assets are sooner specific to a single transaction partner. Due to their small size and corresponding effects of scale and scope, small firms typically cannot aim for a low cost strategy in a large market for commodities, but need to go for a differentiation and/or focus strategy, in Porter's terms, or for a product innovation. Thus they will sooner get into the situation of a specialized product requiring specialized assets, for a narrow group of customers.

A fourth problem lies in the provision of safeguards by the buyer, to reduce the risk of specific assets to the small scale producer. If a buyer yields a safeguard (ownership of specific assets, guaranteed volume and price of purchase), he will want to control for misuse. The problem is that there are threshold costs in setting up these controls, and that due to lack of systematic and reliable, formal information systems or records in the smaller firm, it is more difficult to design and implement the controls. For example, if the buyer is to own the assets, he will want to make sure that they are not used to supply his competitors.⁸ If he participates in the risk of price increases of inputs for the supplier, or of disruptions in their supply, he will want to be able to verify them. If he guarantees a severance payment under certain specified conditions, he wants to be able to check the conditions. This checking of conditions is difficult when reliable records are lacking on the part of the producer. The problem does not disappear if we go from bilateral to trilateral governance, because then the arbitrator is confronted with these higher

costs of control. An implication for the policy of small producers might then be as follows: one may need formal administrative and monitoring procedures, not for internal management but as a condition for receiving safeguards from customers for transaction specific investments.

Empirically, Semlinger (1991a, b) reports that in fact small suppliers to the car industry supply products with highly transaction specific assets (moulds, dies), without obtaining, and in some cases without even wanting, the safeguards that TCE prescribes. In particular, suppliers do not want the buyer to own the transaction specific assets. How can this be explained? We will consider this when we turn to implications for policy at the end of the article.

Users also may have to invest in transaction specific assets, such as knowledge about the supplier and his product, adjustments in processes to receive and incorporate the product, or systems for monitoring the supplier. Furthermore, in case of transaction specificity on the part of the supplier, the user runs a risk of discontinuity in supply, quality or price of the product used. The main problems for the small user are the generic problems already discussed (bounded rationality, risk, and more opportunism and volatility perceived by transaction partners). A small user may be less able to monitor transaction partners and to find and evaluate alternative sources of supply. He may also lack a sufficient administrative basis for allowing transaction partners, or a third party arbitrator, to exercise monitoring on him.

A problem that weighs particularly heavily for small users is the problem of tacit knowledge. For reasons already discussed, knowledge is often more tacit in smaller firms. A positive side to this may be a more intuitive and daring type of conduct. A negative side of it is that there are greater problems of becoming aware of limitations in present activities and of opportunities for improvement. According to Rogers (1983), the adoption of innovations occurs along stages of "knowledge, persuasion, decision, implementation and confirmation". Awareness is the crux in the first stage of knowledge, but awareness is problematic to the extent that knowledge is tacit, and knowledge tends to be more tacit in small firms. In other words, small firms tend to take current methods more for granted, and are less susceptible

to awareness of needs and opportunities for employing new technology. Their suppliers therefore have to take more trouble in generating such awareness, which raises their costs of contact (marketing) that were already relatively high due to threshold costs. In subsequent stages of adoption there are also effects of firm size, related to the other problems of bounded rationality discussed before: lack of internal expertise to conduct competent search, evaluation and implementation.

9. Implications for policy

Apart from a number of details that are to be settled empirically, the analysis of effects of firm size on transaction costs clearly indicates that on the whole transaction costs are higher for the smaller firm. They are not higher to the same extent for all firms, in view of the diversity of SME, but the effect is systematic and pervasive. Costs of governance schemes to reduce transaction costs also are often relatively higher for small firms: transactions may be too small to be worth the bother of such a scheme. As a result, *ceteris paribus* transactions will gravitate to larger partners. In case only small suppliers are available, users will be more inclined to make rather than buy, or there will be a preference for trilateral rather than bilateral governance. In vertical relations between large and small firms, the larger partner will often face lower transaction costs. Such asymmetry of transaction costs leads to a one-sided dependence of the smaller partner, in case he cannot develop countervailing power. The problem can be mitigated in several ways by small firms and by public policy.

Individually, of course small firms can try to compensate for transactional weaknesses by mobilizing the typical behavioural strengths of small business identified in the literature: market flexibility, motivation, entrepreneurial drive, unorthodox ventures, niche markets, being "lean and mean". They may have to find compensation, in particular, in the offer of a unique product or competence. The competence may entail that one acts as a signalling and triggering device for transaction partners, to compensate for obstacles in their perception and evaluation of opportunities and threats. In other words, when we take into account processes of learning, a new dimension is

added to the role of transaction relations. As a transaction partner one may play such a pivotal role in generating complementary perceptions and competences, that one may require no additional safeguards for continuity and reliability of the transaction relation. Perhaps this contributes to an explanation of the empirical phenomenon reported by Semlinger (1991a, b) that small suppliers do not want buyers to own transaction specific assets of production: that might detract from their position as unique partners in development. Alternatively, when the small partner has a unique offering, he may be in a position to demand other measures to achieve a more balanced dependence, even at a relatively high cost of the governance scheme. But perhaps the offering is so unique and valuable that symmetry of dependence is thereby already achieved. Alternatively, a small producer can try to reduce transaction specificity of assets by achieving product differentiation by adding on features to a standard product, or by flexible assembly of standard components, rather than by building uniqueness in with transaction specific machinery, skills, components or other inputs, thus lessening dependence on the transaction partner and the corresponding need of expensive governance schemes. Lack of capabilities in the collection of external information, or in the selection and governance of specialist suppliers of business services (in second order transactions needed to govern first order transactions), may be mitigated by the cunning use of largely informal personal networks of the entrepreneur. In these networks social or personal bonds of trust may reduce risks of opportunism.

Collectively, small firms can try to achieve advantages by cooperation. These may lie in joint economies of scale, scope or knowledge in production or transaction costs. Or they may lie in a unique combined offering to strengthen the uniqueness of their product, thus creating a better balance of dependence between large and small transaction partners. Such a unique combined offering might be a sub-assembly of components, an integration of previously separate components, or a unique product-service combination (product plus capability for R&D, logistics, installation, training, maintenance and repairs, etc.). This is the force of the arrangement of "industrial districts".

Government, political parties or small firms

collectively can try to reduce transaction costs or the cost of governance schemes by furthering norms that restrict opportunistic business conduct; by stabilizing market conditions to reduce exogenous uncertainty; by furthering technical standards or standardized procedures that reduce in particular the threshold costs of search, evaluation, contracting, monitoring and arbitration. They can also strive to maintain or improve norms or standards of ethical conduct more in general. Note that there is a double effect involved. There is a reduction of transaction costs for small firms themselves, and a reduction of their scope for opportunism, which reduces transaction costs for their transaction partners. This is in line with the argument of North (1990) that the role of institutions is to reduce transaction costs and thereby enable the specialization on which prosperity depends. In view of the relatively high transaction costs in SME, it is in particular SME that has a stake in those institutions. The institutions need not all be set up by central government. Small business itself may set up associations that supply monitoring of their members, or guarantees against malpractice, unethical conduct, inferior quality, business failure, in order to reduce risks for transaction partners. However, such arrangements generate problems of collective action, which we cannot treat adequately in the present article.⁹

To the extent that there is a lack, or failure, of such institutions, small firm managers will have to spend more time on the development and maintenance of external networks, and the networks will be based more on informal, personal ties. This might explain the empirical observation that in Italy significantly more time appears to be spent on setting up and maintaining contacts than in Sweden, Northern Ireland and the USA (cf. Birley, Cromie and Myers, 1991). One would expect this effect to be even stronger in underdeveloped countries.

Problems of adoption on the part of small users, due to tacit knowledge, high search costs and lack of technical expertise may be mitigated by government-sponsored agencies for technology transfer or other facilities for advice or training. In general, such services serve to compensate for thresholds in transaction costs.

Notes

¹ Williamson's assumption here, made in a footnote (1985, p. 92), is that a user cannot produce the input for the total market, thus achieving equal scale, because other users will not be willing to purchase from a competitor. On the whole this assumption is plausible, but there may well be exceptions. When assets are not very specific to the product, users do not run high risks in buying it from a competitor. In fact, one does observe mutual supply of inputs between competitors, as in the aircraft and car industries.

² Williamson (1985, pp. 143–144) himself admitted this shortcoming of TCE: "... the study of economic organisation in a regime of rapid innovation poses much more difficult issues than those addressed here" and "New hybrid forms of organisation may appear in response to such a condition ... Much more study of the relations between organisation and innovation is needed". The development of a more dynamic theory of transaction costs is under way. Preliminary results have been presented in Nooteboom (1992) and Berger, Noorderhaven, Nooteboom and Pennink (1993).

³ For a preliminary analysis, see Berger, Noorderhaven, Nooteboom and Pennink (1993) and Nooteboom (1992).

⁴ The term "threshold costs" was introduced in studies of economies of scale in retailing and other services by Nooteboom (1982).

⁵ Cf. the sources mentioned above.

⁶ For a discussion of such problems, see e.g. Elster (1989). For an empirical survey of the problem in the context of developing standards for services to small business see Nooteboom, Zwart & Bijmolt (1992).

⁷ Also, considering problems of collective action, one may doubt the credibility of such safeguards. If the officials of the trade organization depend for their position on the support of the members, they may lack the incentive to design and enforce severe monitoring and sanctions. Officials and members may have a joint incentive towards "window dressing", with safeguards and sanctions that look good but in fact have little bite. In due course, knowledgeable transaction partners will know this, and the expense on the trade association may have little real benefit.

⁸ Of course if those assets are believed by the buyer to be fully transaction specific, there is no such opportunity for the producer, but the buyer may not be certain about the degree of specificity.

⁹ For a treatment of problems of collective action, see Elster (1989).

References

- Berger, J., N. G. Noorderhaven, B. Nooteboom, and B. J. Pennink, 1993, 'Understanding the Subcontracting Relationship: The Limitations of Transaction Cost Economics', in J. Child, M. Crozier, R. Mayntz et al., *Societal Change Between Market and Organization*, Frankfurt: Campus/Westview, in press.
- Birley, S., S. Cromie, and A. Myers, 1991, 'Entrepreneurial Networks: Their Emergence in Ireland and Overseas', *International Small Business Journal* 9(4), 56–74.
- Bol, A., A. Nijssen, G. Regter, and F. van Rossum, 1982, *Employment and Labour Volume in SME, 1960–1980* (in Dutch), Zoetermeer, NL: EIM.
- Coase, R. H., 1937, 'The Nature of the Firm', *Economics N.S.* 4, 386–405.
- Elster, J., 1989, *The Cement of Society: A Study of Social Order*, Cambridge: Cambridge University Press.
- Granovetter, M., 1982, 'The Strength of Weak Ties', in P. Marsden and N. Lin, *Social Structure and Network Analysis*, Beverly Hills: Sage, pp. 105–130.
- Johannisson, B., 1986, 'Network Strategies, Management Technology for Entrepreneurship and Change', *International Small Business Journal* 5(1), 20.
- Johnson, J. L. and R. Kuehn, 1987, 'The Small Business Owner/Manager's Search for External Information', *Journal of Small Business Management* 24, 53–60.
- NOBIN, 1984, *External Information Provision of the Small and Medium Sized Manufacturing Firm* (in Dutch), The Hague: NOBIN.
- Nooteboom, B., 1982, 'A New Theory of Retailing Costs', *European Economic Review* 17, 163–186.
- Nooteboom, B., 1992, 'Towards a Dynamic Theory of Transactions', *Evolutionary Economics* (forthcoming).
- Nooteboom, B., 1993, 'An Analysis of Specificity in Transaction Cost Economics', *Organization Studies* (forthcoming).
- Nooteboom, B., P. S. Zwart, and T. Bijmolt, 1992, 'Transaction Costs and Standardisation in Professional Services to Small Business', *Small Business Economics* 4(2), 141–151.
- North, D. C., 1990, *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
- Polanyi, M. 1962, *Personal Knowledge*, London: Routledge.
- Rogers, E. M., 1983, *Diffusion of Innovations*, New York: MacMillan.
- Rothwell, R., 1985, 'Innovation and the Smaller Firm', First International Technical Innovation and Entrepreneurship Symposium, Utah Innovation Foundation, Salt Lake City, 11–13 September.
- Semlinger, K., 1991a, 'New Developments in Subcontracting – Mixing Market and Hierarchy', in A. Amin and M. Dietrich (eds.), *Towards a New Europe: Structural Change in the European Economy*, Aldershot: Edward Elgar.
- Semlinger, K., 1991b, 'Innovation, Cooperation and Strategic Contracting', paper presented at the international colloquium on 'Management of Technology: Implications for Enterprise Management and Public Policy', Paris, 27–28 May.
- Smeltzer, L. and G. Fann, 1989, 'Communication Patterns in Large, Bureaucratic Organizations Compared to Small Entrepreneurial Organizations', *Group and Organization Studies*, 198–215.
- Williamson, O. E., 1975, *Markets and Hierarchies: Analysis and Antitrust Implications*, New York: The Free Press.
- Williamson, O. E., 1985, *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, New York: The Free Press.

Copyright of Small Business Economics is the property of Kluwer Academic Publishing / Business and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.