The Comparative Organisation of Large and Small Firms: An Information Cost Approach

Mark Casson

ABSTRACT. The paper introduces the concept of information cost and explains its relevance to the organisation of the firm. A distinction is drawn between autocratic and consultative management styles, and the economic logic of each style is explained in terms of information costs. It is argued that autocratic management is a rational organisational response to a business environment that contains a single dominant source of volatility, while consultative management is a rational response to multiple sources of volatility. It is argued that small firms exploit the advantages of the autocratic style. Small firms are therefore to be found predominantly in industries dominated by a single source of volatility.

1. Introduction

This paper develops a general model within which the organisational strategies of large and small firms can be compared. The usual approach to organisational issues these days is through the theory of transaction costs (Coase, 1937; Williamson, 1985). Transaction costs are certainly a useful way of explaining the boundaries of the firm: large firms or small firms are created according to whether the boundaries are widely or narrowly drawn. However, transaction cost theory is not so successful in explaining what goes on inside these boundaries.

This is because transaction cost theory has paid insufficient attention to information costs. Information costs, in this context, are the costs of using information in decision-making. Many of the costs of running an organisation are, in fact, information costs rather than transaction costs. They include, for example, costs incurred in observing the environment, communicating the observations, and substituting the observations into a decision rule. Information costs are related to the *quantity* of information that is collected and the number of stages it must pass through before it is used to make a decision. Transaction costs, by contrast, are concerned principally with the *quality* of information. They are the costs of assuring that people do not supply wrong information in order to mislead other people.

As the complexity of the information flows handled by a firm increases, both information costs and transaction costs increase. While owner's efforts to control transaction costs are the major determinant of the boundaries of the firm, the control of information costs is the major determinant of what goes on inside. It is these information costs which are the focus of this paper. Transaction costs are considered too, but only in relation to information costs, and only in the specific context of the setting of budgets within the firm.

2. Volatility

The main thrust of the paper is to show that organisational structure can be explained as the outcome of attempts to minimise information costs. This is the analogue, in a way, of explaining the boundaries of the firm in terms of the minimisation of transaction costs. Indeed the operationalisation of information cost theory is in some ways more straightforward than is the operationalisation of transaction cost theory. In particular, the volatility of the firm's environment turns out to be a major determinant of information costs, and a major

Final version accepted on August 18, 1995

Department of Economics
The University of Reading
P.O. Box 218
Reading, RG6 6AA,
U.K.

Small Business Economics 8: 329–345, 1996.

influence on the organisational strategies used to reduce them (Casson, 1994b).

If this theory is to be relevant to small firm economics then it must explain some of the "stylised facts" revealed by empirical research. There is considerable anecdotal evidence, for example, that small manufacturing firms adopt a more autocratic style of management than do large ones. An autocratic style signifies a reluctance to consult other people before taking a decision. As the theory shows, autocracy reflects an implicit belief by the autocrat that it is he himself, and no one else, who possesses the key information relevant to the decision.

Small firms also seem to be more flexible than large ones in accommodating sudden changes in demand because they are more able and willing to vary capacity utilisation. Small firms rely less on sophisticated routine procedures than do large firms, but some of them appear to compensate for this by a face-to-face culture in which charismatic leadership substitutes for impersonal control mechanisms. This can also be explained by the general theory, as demonstrated in the second part of the paper.

The crucial link between theory and evidence is effected by identifying small firms as the firms that specialise in operating in volatile environments. More particularly, they operate in environments which have a single major source of volatility. *A single major source of volatility, theory shows, supports the very kind of autocratic management style that seems to be associated with small firms.* Large firms, by contrast, tend to operate in relatively stable environments. More precisely, they operate in environments where there are multiple sources of volatility, but where no one source of volatility is sufficiently large to dominate the others. Such environments favour a consultative management style which is associated with relatively complex decision-making procedures. This management style is the one that tends to be adopted by larger firms.

On this view, small firms remain small because, given the nature of their environment, they do not require a large management team. Large firms are large because the nature of their environment requires extensive consultation within a team composed of different specialists. Large firms may also become large because their decision procedures evolve into proprietary assets which can be exploited by taking over the management of other firms as well. Small firms do not have this potential because their management skills are embodied entirely in the owner-manager of the firm.

Because different industries experience different patterns of volatility, some industries are suited to autocratic management and others to consultative management. Industries suited to autocratic management will tend to consist of small firms whilst industries suited to consultative management will tend to be dominated by a small number of large firms.

The pattern of volatility may also vary over the lifetime of the industry, with volatility in general being particularly high during the initial innovative phase and again during the terminal phase in which the industry is threatened by obsolescence caused by innovations elsewhere. Small firms should therefore be favoured in both the early and late stages of the industry life cycle. Such organisational imperatives may prove difficult to reconcile with technological imperatives, however. In the final analysis it will be the interplay of organisational and technological factors which determines how the size of the firm varies over the life cycle in any given industry.

At its most basic, an autocratic style of management does little more than refer every major decision to the "boss". The essence of autocracy is that the boss makes the decision without consultation. Other people may be employed to implement his decisions, but they do not have much opportunity to advise him on what to do. The decision-making procedures of an autocratic firm are therefore very simple ones. By contrast, decision-making in a consultative firm normally involves quite complex procedures.

Some of these procedures may be formalised, whilst others may be based on an accumulated set of precedents instead. They are likely to be finely tuned to the firm's environment, and therefore specific to the industry – perhaps, indeed, specific to the firm itself. These procedures are transmitted from one generation of employees to the next by special in-house training, as well as by socialisation within the management team.

In small firms, by contrast, the important features of the firm's environment are not encoded in procedures but simply "kept inside the boss's

head". The few procedures that are employed will tend to be derived from imitation of what other small firms in the same industry or the same area do. Unless the boss imparts his knowledge of the environment to his successor, the firm may fail to survive from one generation to the next.

3. Decisiveness and autocracy

Autocratic decision-making benefits from access to decisive information. Information is decisive when it is possible to make a rational decision by ignoring all other information. The justification for the autocrat is that whatever other people may know that he does not, knowing it would not alter his decision. In autocratic firms the information that the autocratic claims to possess is often marketing information. This is illustrated in the following model, which is the basis for the analysis in the rest of this paper.

Consider a firm that can source demand from two plants, each of unit capacity. Management has two options for each plant – to utilise it fully or to leave it idle. The issue is to determine what information is required to decide whether to utilise each plant or not – in other words, to determine what information is necessary to formulate a rational output plan.

The firm operates in a stochastic environment. Conditions at each plant can be either good (state 1) or bad (state 2). Conditions vary randomly each period, independently of conditions at the other plant and of conditions at earlier times. The cost of production at the first plant is $c_{11} > 0$ if local conditions are good and $c_{12} > c_{11}$ if local conditions are bad. Similarly, cost of production at the second plant is c_{21} under good conditions, and $c_{22} > c_{21}$ under bad ones. Costs at the second plant are generally higher than those at the first – although not so high that it is not cheaper to produce under good conditions at the second plant than under bad conditions at the first – i.e., $c_{11} < c_{21} < c_{12} < c_{22}$.

Customer demand can be in one of two states – stable or volatile. If it is stable then a unit quantity is demanded all the time. If it is volatile then the quantity demanded is either zero or two, varying randomly according to customers' circumstances.

The maximum price p that customers are willing to pay is always the same. Suppose to begin with the price exceeds the cost of producing under the worst possible circumstances, $p > c_{22}$. It is then easy to show that in the volatile state information on demand is decisive for the output plan, whereas in a stable state it is not.

Consider the volatile state first, beginning with the case of zero demand. Since marginal revenue is zero, whilst marginal costs are always positive, it never pays to produce – so both plants should remain idle. At the opposite extreme, when two units are demanded, it always pays to produce the second unit as well as the first because marginal revenue, p, exceeds the highest possible marginal cost, c_{22} – so that both plants should be used. Thus the use of each plant is governed entirely by demand conditions: information on demand is decisive for the output plan.

Turning to the stable state, the same kind of reasoning indicates that, with a constant unit demand, one (and only one) plant should be used. But demand information alone cannot determine which plant this should be. For with one unit of demand instead of two, the two plants are now substitutes rather than complements for each other. Whilst there is some presumption that the first plant should be used because on average it is cheaper, this is no more than a presumption, because if the first plant faces bad local conditions whilst the second plant enjoys good conditions then the second plant should be used instead. In general therefore, information on both sets of local conditions is required before a rational output plan can be determined. Demand is no longer decisive because information on demand must now be synthesised with information on supply conditions before a decision can be made.

Generally speaking, the more volatile a factor is, the more decisive information about that factor is likely to be. In this context, volatility is measured in terms of the impact of the factor on the value of the decision-maker's objective, such as the profit of the firm. Thus in the stable state fluctuations in cost conditions completely dominate fluctuations in demand, and so there is no way that demand information can be decisive. In the volatile state, on the other hand, fluctuations in demand have a far greater profit impact than do changes in production conditions, and it is this property that renders demand information decisive.

There is, of course, an intermediate case, in which neither demand nor production conditions are decisive. This prevails, for example, when price falls to below the highest possible marginal cost: $c_{12} > p > c_{22}$. In this case production conditions need to be known in order to determine whether to produce a second unit of output when demand is buoyant; conversely, demand conditions need to be known to determine whether to use the second plant as well as the first when conditions at the second plant are good. In the interests of simplicity, however, this intermediate case will not figure in subsequent discussion.

4. Information costs and consultative procedures

It is fairly clear how an autocrat normally proceeds – he consults the source that he regards as providing the decisive information, makes his decision, and instructs others to act accordingly. In the context of the previous model this means that he investigates demand and then instructs the manager of each plant to produce if and only if he discovers that demand is high.

Consultative procedures, however, offer a much wider range of options. Their aim is to synthesise information from different sources. Thus in the stable state, where local conditions at both plants are potentially relevant to the output plan, they serve to combine the information held by each manager. This assumes, however, that the information is worth obtaining. Consultative procedures do not have to involve consulting everyone nor do those who are consulted have to be consulted at the same time. Given these options, then, how does a rational owner of a firm choose between them?

When planning output in a stable environment, six main procedures need to be evaluated. These are listed in Table I. The first procedure, admittedly, is very crude – it simply involves the owner guessing at the state of local conditions and choosing the location that affords the lowest expected cost. The second and third strategies investigate costs at one of the locations but not the other. The owner consults the local manager at the chosen location, but simply guesses the state of the other location himself. The next two strategies are more like conventional procedures in the sense that they involve two steps rather than just one. One of the locations is investigated and if the resulting information supplied by the local manager proves decisive then the investigation

TABLE I
Alternative strategies of synthesising information from different locations to decide where to produce

Number	Nature of strategy
1.	*No investigation.* Estimate the likely conditions at each location and then commit, as appropriate, to either (a) produce at first location, or (b) produce at second location Do not consult local managers.
2.	*Investigate the first location only.* Consult the local manager concerned. If the result of his investigation is decisive (conditions are good) then produce at first location. If the result is indecisive, estimate the likely conditions at the other location and then select (a) or (b) as appropriate.
3.	*Investigate the second location only.* As (2), with the locations interchanged (and bad conditions replacing good).
4.	*Investigate sequentially, beginning with the first location, with an option to stop if a decisive result is obtained.* If the local manager provides decisive information (conditions are good) produce at the first location and investigate no further. If no decisive result is obtained on the first investigation (conditions are bad) consult the manager at the second location. Synthesis then guarantees a decisive result: bad conditions imply (a) and good conditions imply (b).
5.	*Investigate sequentially, beginning with the second location, and with an option to stop if a decisive result is obtained.* As (4), with the locations interchanged (and bad conditions replacing good).
6.	*Investigate both locations.* Consult both managers at the outset. With full information, select (b) if conditions are bad at the first location but good at the second, and (a) otherwise.

stops and a decision is made. If the information is not decisive then the other location is investigated as well. The final strategy involves investigating both conditions at the outset, without waiting to see whether one of them will generate decisive results. Thus both of the local managers are consulted from the start under this procedure.

Now it goes without saying that information, provided it is accurate, generally improves the quality of a decision. Certainly it cannot make a rational decision any worse. If information were costless, therefore, it would always pay to collect all the information available. It would not matter if some of the information were irrelevant, because no resources would have been wasted in collecting it. This implies that the final strategy would normally be preferred in the absence of information costs.

In practice, of course, information is costly to obtain. In the context of the present model, this means that it is costly for the local manager to investigate conditions and to communicate the results to the owner of the firm. Such information costs have two important implications.

The first is that irrelevant information should no longer be collected (unless its collection cross-subsidises another activity). This has a bearing on the viability of the final strategy, because this strategy always collects a second item of information whether the first item is decisive or not. With costly information this strategy is normally inferior to sequential strategies which terminate the search for information once decisive information has been found. The only qualification is that simultaneous investigation may sometimes be cheaper than sequential investigation because of economies of scope, so that one investigation can subsidise another. Such cases seem uncommon in practice, though. Indeed, because of limited managerial capacity, investigations may have to be carried out sequentially in any case.

Which of the two sequential strategies should be chosen depends upon the probabilities with which they generate decisive information at the first stage, and on the relative costs of investigating the two local sources of information. In this context decisive information is information that shows either conditions at the first plant to be good or conditions at the second plant to be bad – in either case the implication is that the first plant and not the second plant should be used. It follows that if conditions are more likely to be good at the first plant (where costs, on average, are lower) than to be bad at the second plant (where costs, on average, are higher) then strategy 4, which investigates conditions at the first plant first, is the more likely to provide decisive information. Conversely, if conditions are more likely to be bad at the first plant than they are good at the second plant then strategy 5, which investigates conditions at the second plant first, is the better bet.

The main qualification arises when the local costs of investigation differ. If investigation is cheaper at the first plant than the second then it may pay to investigate the first plant first, even though it is less likely to afford decisive information, because the potential saving of investigation costs at the second stage is greater. This strategy also defers the more expensive investigation until later, so the present value of investigation costs is reduced even if the second stage investigation does have to be carried out.

The second implication of information costs is even more profound. When information is costly to obtain, decision-makers may opt to do without some of the information that they really require. They face a trade-off between the expected value of the outcome of the decision and the information cost incurred. When the expected value of the information, in terms of the decision-maker's objective, is less than the cost of acquiring it, a rational decision-maker will do without the information even at the risk of making a wrong decision. In the presence of information costs, therefore, it is often rational to behave in what may appear to an observer to be a "boundedly rational" way (Casson, 1994a).

It is, therefore, potentially misleading to interpret bounded rationality in terms of, say, managerial naivety, or as a consequence of non-rational programmed behaviour. The use of boundedly rational decision rules may, on the contrary, involve a complex trade off involving sophisticated economies of information cost. This can be seen in the way that strategies 2 and 3, involving a single investigation, might be substituted for the sequential strategies 4 and 5 as a rational response to information cost.

Under strategy 4 the investigation of conditions

at the first plant is followed by an investigation of conditions at the second plant if conditions at the first plant turn out to be bad. This is only reasonable, it would seem, for if the "first choice" plant has bad conditions then the "second choice" plant should be checked out so that if conditions there are good then it can be used instead. But suppose that conditions at the second plant are almost certainly good. Is it worth investigating conditions when one is subjectively almost certain what the result will be? Certainly not, provided tha. .he cost of investigation is high and the consequences for profit of an (occasional) error are not too severe. The second stage of strategy 4 is worth proceeding with only when subjective uncertainty is reasonably high, investigation costs are reasonably low, and the consequences of producing at the second location under bad conditions are reasonably severe.

Indeed, if these conditions are not satisfied in respect of conditions at the first location, then even strategy 2 may be abandoned in favour of strategy 1. If, for example, the manager is subjectively certain that conditions are good at the first plant and bad at the second then he may as well go ahead and commit to production at the first plant rather than incur costs finding out what he believes he already knows. Conversely, if he is subjectively certain that conditions are bad at the first plant but good at the second then he may commit to production at the second plant without any investigation. These two commitments represent the alternative versions of strategy 1 identified in the table.

These ideas may be formalised as follows. Let $b_1, b_2 \geq 0$ be the respective costs of investigating local conditions one at a time and let p_1, p_2 $(0 \leq p_1, p_2 \leq 1)$ be the subjective probabilities of good local conditions (i.e., that state 1, as defined earlier, prevails locally). Simultaneous investigation affords a discount $a \geq 0$. In a stable environment, as assumed above, the output strategy $i = 1, 2$ is simply identified by the plant which is used. The profit generated by output strategy i when plant i is in state j is then $\pi_{ij} = p - c_{ij}$ $(i, j = 1, 2)$ where c_{ij} is the production cost defined earlier.

Consultative strategies are indexed by $k = 1, \ldots, 6$, as above. A rational risk-neutral owner of a firm operating in a perfect market environment maximises expected profit, v. Profit is here measured net of expected information cost as well as expected production cost. Let v_k be the expected profit generated by an optimal output plan used in conjunction with consultative strategy k. The owner chooses k to maximise v.

The values of the various consultative strategies are summarised in Table II. These values are all linear functions of the probabilities p_1, p_2, the investigation costs a, b_1, b_2, and the profit parameters π_{11}, π_{12}, π_{21}, π_{22}. The strategies v_1, v_2 and v_3 each have two functional forms, however, which apply to the different output plans associated with different combinations of probabilities p_1, p_2. In each case there is some critical probability where, in the absence of decisive information, the expected values of the alternative output strategies are equal. Each critical probability marks a switch from one output plan to another.

These critical probabilities p_1^*, p_2^* and p_3^* divide the probability space into five regimes, as indicated in Figure 1. Within each regime the value of each strategy has a fixed functional form. This makes it easy to ascertain which strategy has the highest value in each regime.

A typical set of results is illustrated in Figure 2. As explained in the note to the table, these

TABLE II

Expected values of the consultative strategies identified in Table I

$$v_1 = \begin{cases} v_1^+ = \pi_{12} + p_1(\pi_{11} - \pi_{12}) & \text{If } p_1 \geq p_1^*(p_2) \\ v_1^{+-} = \pi_{22} + p_2(\pi_{21} - \pi_{22}) & \text{If } p_1 < p_1^*(p_2) \end{cases}$$

where

$$p_{11}^* = ((\pi_{21} - \pi_{22})p_2 - (\pi_{12} - \pi_{22}))/(\pi_{11} - \pi_{12})$$

$$v_2 = \begin{cases} v_2^+ = p_1\pi_{11} + (1 - p_1)\pi_{12} - b_1 & \text{If } p_2 \leq p_2^* \\ v_2^{+-} = p_1\pi_{11} + (1 - p_1)(\pi_{22} + p_2(\pi_{21} - \pi_{22})) - b_1 & \text{If } p_2 > p_2^* \end{cases}$$

where

$$p_{22}^* = (\pi_{12} - \pi_{22})/(\pi_{21} - \pi_{22})$$

$$v_3 = \begin{cases} v_3^+ = p_1\pi_{11} + (1 - p_1)\pi_{12} - b_2 & \text{If } p_1 \geq p_3^* \\ v_3^{+-} = p_2\pi_{21} + (1 - p_2)(\pi_{12} + p_1(\pi_{11} - \pi_{12})) - b_1 & \text{If } p_1 < p_3^* \end{cases}$$

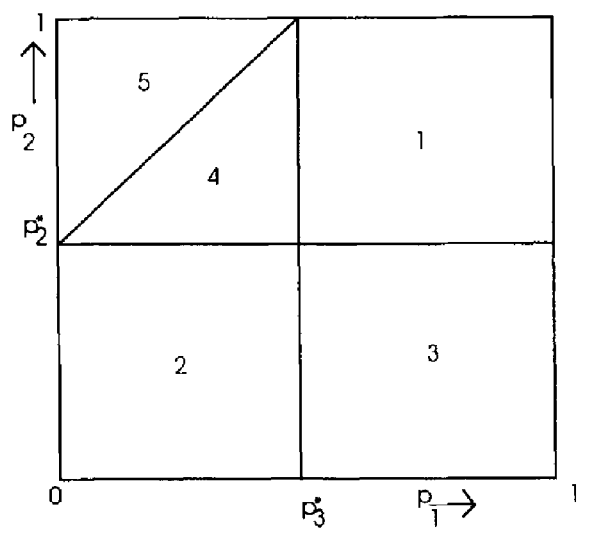
where

$$p_{13}^* = (\pi_{21} - \pi_{12})/(\pi_{11} - \pi_{12})$$

$$v_4 = p_1\pi_{11} + (1 - p_1)(\pi_{12} + p_2(\pi_{21} - \pi_{12})) - b_1 - (1 - p_1)b_2$$

$$v_5 = p_1\pi_{11} + (1 - p_1)(\pi_{12} + p_2(\pi_{21} - \pi_{12})) - b_2 - p_2b_1$$

$$v_6 = p_1\pi_{11} + (1 - p_1)(\pi_{12} + p_2(\pi_{21} - \pi_{12})) - b_1 - b_2 + a$$



Notes:

1. The regimes are identified by the squares and triangular areas labelled 1–5. Let v_{ij} be the value of consultative strategy i in regime j ($i = 1, \ldots, 6$; $j = 1, \ldots, 5$). Then from Table 2:
   $v_{11} = v_{12} = v_{13} = v_{14} = v_1^+$; $v_{15} = v_1^{++}$
   $v_{22} = v_{23} = v_2^+$; $v_{21} = v_{24} = v_{25} = v_2^{++}$
   $v_{31} = v_{33} = v_3^+$; $v_{32} = v_{34} = v_{35} = v_3^{++}$
   $v_{4j} = v_4$; $v_{5j} = v_5$; $v_{6j} = v_6$ ($j = 1, \ldots, 5$)
2. The case illustrated in the figure corresponds to $\pi_{11} = 4$, $\pi_{21} = 4$, $\pi_{12} = 4$, $\pi_{22} = 4$, whence $p_2^* = p_3^* = 0.5$.

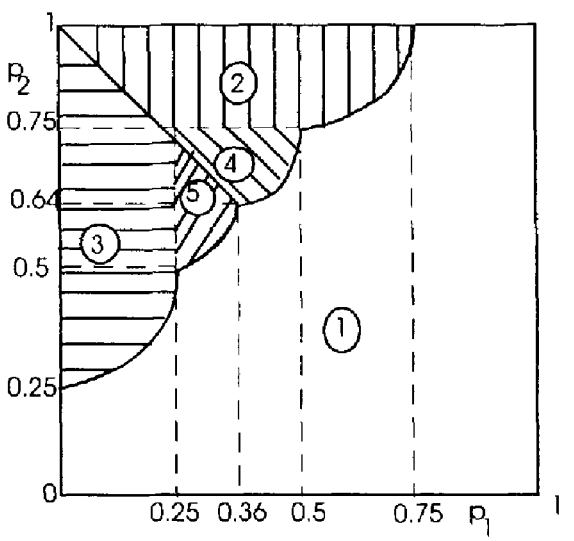
Fig. 1. Five regimes for evaluating consultative strategies.

results are generated by a case in which investigation costs are fairly significant and are the same for each location. The optimal consultative strategy is identified by the circled number in the relevant region of the figure. It can be seen that five out of the six possible consultative strategies are used under some conditions. Only the simultaneous investigation strategy is not used, because in this particular case it affords no economy in investigation costs.

Because investigation costs are significant, the "no investigation" strategy is used whenever the first location seems the obvious choice – i.e., when conditions there are likely to be good or conditions at the other location are likely to be bad. It is only when bad conditions at the first location are likely to be combined with good conditions at the other location that any kind of investigation is likely to be worthwhile.

Because the investigation costs are equal, the figure is symmetrical about a diagonal running down from left to right. Opposite sides of the diagonal simply interchange which location is investigated first. Sequential investigation is adopted only when subjective certainty about conditions at both locations is relatively low. This explains why the regions 4 and 5 are in the interior of the figure and regions 2 and 3 along its boundaries, rather than the other way round.

The derivation of such results is straightforward, but tedious. Solution of the model becomes even more tedious if some of the earlier assumptions are relaxed. For example, it has been assumed that the costs of the two locations under good and bad conditions "interleave", in the sense that the cost of the first location under bad conditions lies between the costs associated with the second location: $c_{21} < c_{12} < c_{22}$. Other cost structures need to be considered to give a complete account of the model, however. While some of these alternatives are trivial, in the sense that one location dominates another whatever conditions prevail, there is a non-trivial alternative in which one location is effectively "high risk" and the



Notes:
The optimal strategy for each combination of probabilities is indicated by the circled number. The calculations have been made for the values $\pi_{11} = 4$, $\pi_{21} = 3$, $\pi_{12} = 2$, $\pi_{22} = 1$. $a = 0$, $b_1 = b = b_2 = 0.5$.

Fig. 2. Influence of probabilities on optimal consultative strategy.

other "low risk". An example is where the costs of the second location are nested between the costs of the first, so that the previous inequality is partially reversed: $c_{21} < c_{22} < c_{12}$. This means that the result of investigating the high risk location is always decisive for the output plan. It can then be shown that when the costs of investigating each location are equal, as assumed in Figure 2, only strategies 1 and 2 will ever be used.

The existence of alternative cost structures is not the only complication that can arise. It has been assumed throughout that each of the factors being investigated affects only one of the strategies. This is reflected in the concept of investigating the local conditions relating to a given plant. But in a different context both factors may affect both strategies (as considered in Casson, 1994b). This means that unlike the present example single investigation may never produce a decisive result at all. This argues against single step investigation strategies such as 2 and 3, and brings the sequential strategies 4 and 5 much more to the fore. Unfortunately, it renders the evaluation of these strategies more complicated at the same time.

5. Organisational memory

So far no attention has been paid to the complexity of the consultative strategies, and more particularly to the impact of this complexity on the cost of devising and memorising the procedures they involve. In the context of the present model the consultative strategies and the procedures associated with them represent the fundamental building-blocks of the organisation – the recurrent implementation of these procedures, period by period, is what the organisation exists to do.

This section considers the strategies that a large firm can adopt to transmit its procedures from one generation of employees to the next. A particular concern is how the optimal procedure for the stable state is memorised during a volatile period in which it is over-riden by autocratic behaviour. This issue is particularly relevant to the debate over "flexible organisations", for it explains how in an evolving environment, the firm can alternate between "large firm" and "small firm" modes of operation by retaining its "large firm" skills during a period of "small firm" operation.

The cost of memorising procedures, it is assumed, depends upon two main factors: the complexity of the procedure m and the personal qualities of the members of the management team, n. An index of complexity should reflect, roughly speaking, the number of instructions that need to be memorised. On this criterion the "no investigation" strategy 1 is the simplest whilst the sequential strategies 4 and 5 are the most complex, with the single-investigation strategies 2 and 3 somewhere in between. It is a little more difficult to rate the simultaneous investigation strategy 6. Since, however, it is no more difficult, in principle, to remember to investigate both locations than it is to remember which one to investigate, it could be said to have the same complexity as strategies 2 and 3. If this argument is accepted then it is appropriate to distinguish just three levels of complexity.

Thus if m_k is the complexity of consultative strategy k it is appropriate to specify the inequalities $m_1 < m_2 = m_3 = m_6 < m_4 = m_5$. Further simplification can be effected by assuming that it is completely trivial to remember the autocratic strategy 1, so that $m_1 = 0$. The total cost of memorising strategy k for one period can then be expressed as $m_k n$, where n measures the scarcity of intellectual qualities amongst the management team.

It is easy to memorise a procedure when it is in use but difficult to memorise it when it is not. Consultative procedures are not in use when the volatile environment prevails. It is therefore assumed that memorising procedures is costless when the stable state prevails and that memory costs are incurred only when the volatile state prevails instead.

An obvious way to economise on memory costs is to re-invent the consultative procedure each time stable conditions recur. Such repeated improvisation incurs costs of its own, however. Thus the owner faces a trade-off between memory costs and re-invention costs. The way this trade-off is optimised will determine the level of costs that an efficient firm incurs on account of its procedural complexity.

There is another issue related to this, which concerns whether or not the owner of the firm should defer his choice of the consultative procedure until the stable state actually occurs. If the

firm commences operations while demand is in a volatile state, then the owner can devise his strategy for the stable state in advance, thereby avoiding hasty improvisation later on. On the other hand, he then has to memorise his strategy until the stable state actually occurs. Thus an initial decision saves costs of improvisation but at the expense of greater memory cost at the outset.

The situation may be analysed formally by supposing that the cost of devising the optimal procedure is some multiple of the cost per period of memorising it later. This multiple is $h_1 > 1$ if the procedure is devised at the outset and $h_2 > h_1$ if it is improvised later on.

Stable demand conditions and volatile demand conditions occur in intermittent spells. The transience of a given state of demand is measured by the probability that demand will switch to the other state at the start of a given period. The transience of the stable state is θ_1 and the transience of the volatile state is θ_2 ($0 \leq \theta_1, \theta_2 \leq 1$). The converse of transience is persistence. Thus the probability that the stable state will persist is $1 - \theta_1$, and the probability that the volatile state will persist is $1 - \theta_2$.

To analyse the deferral of the choice of procedure, it is appropriate to take the volatile state as the initial condition. The probability q_t that the stable state prevails in period t is then

$$q_t = (\theta_2/(\theta_1 + \theta_2))(1 - \theta_1 - \theta_2)^t \qquad (t = 0, 1, 2, \ldots)$$

The probability that the stable state appears or reappears in period t after a volatile spell is

$$y_t = \begin{cases} 0 & t = 0 \\ \theta_1 (1 - q_{t-1}) & t = 1, 2 \end{cases}$$

and the probability that the stable state occurs for the first time in period t is

$$z_t = \begin{cases} 0 & t = 0 \\ \theta_2 (1 - \theta_2)^{t-1} & t = 1, 2 \end{cases}$$

The selection of an optimal procedure must now be regarded as an investment strategy, and so it is appropriate to apply a discount rate $r > 0$ to future profit flows. A corresponding change is required in the owner's objective: he now maximises wealth rather than profit, where wealth is measured by the expected net present value of the profit stream.

The wealth generated by alternative memory strategies applied to the kth procedure is indicated in Table III. The values shown exclude profits generated in the volatile state, since these are independent of the procedure used in the stable state. The table gives a general expression for the value and then shows the formula obtained by substituting in the algebraic equations above.

The value of a decision at the outset, w_{1k}, is arrived at by deducting both the initial cost of the decision and the intermittent cost of memorising the procedure through volatile spells of demand thereafter. The value of setting a precedent by improvising a decision the first time that stable conditions occur is given by w_{2k}. Its value reflects a higher decision cost – albeit a discounted one, because the decision is deferred. On the other hand, it involves lower memory costs, because memory is not required until after the first stable

TABLE IV
Paired comparisons of alternative memory strategies

$$w_{1k} = \sum_{t=0}^{\infty} v_k q_t/(1 + r)^t - \sum_{t=0}^{\infty} m_k n(1 - q_t)/(1 + r)^t - h_1 m_k n$$
$$= (1 + (1/r))(((v_k + m_k n)\theta_2/(r + \theta_1 + \theta_2)) - m_k n) - h_1 m_k n$$

$$w_{2k} = \sum_{t=0}^{\infty} v_k q_t/(1 + r)^t - \sum_{t=0}^{\infty} \sum_{s=0}^{t} m_k n z_s (1 - q_{t-s})/(1 + r)^t - \sum_{t=0}^{\infty} h_2 m_k n z_t (1 + r)^t$$
$$= (\theta_2/(r + \theta_2))((1 + (1/r))(((v_k + m_k n)(r + \theta_2)/(r + \theta_1 + \theta_2)) - m_k n) - h_2 m_k n)$$

$$w_{3k} = \sum_{t=0}^{\infty} v_k q_t/(1 + r)^t - \sum_{t=0}^{\infty} h_2 m_k n y_t/(1 + r)^t$$
$$= (\theta_2/r)(((v_k(1 + r)h_2 m_k n\theta_2)/(r + \theta_1 + \theta_2)) - h_2 m_k n)$$

spell has come to an end. Improvising a decision without setting a precedent is valued at w_{3k}; while it avoids memory costs altogether, it involves the additional expense of improvising a decision again every time the stable state recurs.

It is obvious from these formulae that a scarcity of intellectual qualities raises the costs of the more complex consultative procedures. Thus where a management team lacks intellectual qualities it will tend to substitute simple procedures for more complex ones, and will especially favour the autocratic procedure of discarding production information altogether. Since the costs of adopting this procedure are lower in the volatile state than in the steady state, it suggests that managers with low intellectual qualities will concentrate on managing in volatile environments where sophisticated procedures are not required. It should be emphasised that these intellectual qualities refer only to the selection and memorising of procedures. They are not a reflection of the general entrepreneurial qualities of the individuals concerned. For as noted in Section 7, volatile environments require considerable skill in diagnosing difficult symptoms. It is rather that volatile and stable environments require different kinds of intellectual ability, with stable environments favouring the more abstract analytical skills and memory skills inculcated by formal education.

The formulae also show that persistence of stable conditions (i.e., low θ_1) encourages investment in more complex procedures. This is because the return from the procedures is increased if they tend to be used more often, and the more often they are used the more worthwhile it is to choose the ideal one at the expense of greater complexity.

The memory strategies themselves are compared in Table IV. The key factor here is the probability that the stable state will recur, θ_2. A high probability of recurrence favours choosing strategy at the outset because it means that although the volatile state prevails initially (by the earlier assumption) the stable state is likely to be entered soon. Thus the additional memory costs incurred by the initial decision are likely to be low. A high probability of recurrence also favours the setting of a precedent when improvising a decision later on. It is clearly expensive to keep improvising if the same situation keeps recurring. Thus an increase in the probability of recurrence tends to shift the owner first from ordinary improvisation to the improvisation of a precedent, and then to making a decision at the outset.

It is fairly obvious that the cost parameters h_1, h_2 will have an impact too. The higher the cost of making a decision relative to the cost of remembering what it was (i.e., the larger are h_1 and h_2), the more advantageous it is to rely on precedent or an initial investigation. The higher the relative cost of improvisation (i.e., the higher is h_2/h_1) the more advantageous initial investigation becomes.

From this analysis the use of precedent emerges as a "middle of the road" strategy, to be employed when the probability of steady state recurrence is modest and the cost advantage to an initial investigation is not too high. This seems to accord well with the "facts of life" in many large firms, where precedents are widely used to deal with issues that

TABLE IV
Paired comparisons of alternative memory strategies

Initial decision versus improvisation of precedent.

$\Delta w_{1k} = w_{1k} - w_{2k} = (((h_2\theta_2 - (1 + r))/(r + \theta_2)) - h_1)m_k n$

$\partial \Delta w_{1k}/\partial h_1 = -m_k n < 0$

$\partial \Delta w_{1k}/\partial h_2 = (\theta_2/(r + \theta_2))m_k n > 0$

$\partial \Delta w_{1k}/\partial \theta_1 = 0$

$\partial \Delta w_{1k}/\partial \theta_2 = (rh_2 + (1 + r))m_k n/(r + \theta_2)^2 > 0$

Improvisation with and without a precedent: A comparison effected at the expected first occurrence of the stable state.

$\Delta w_{2k} = ((r + \theta_2)/\theta_2)(w_{2k} - w_{3k}) = (1 + (1/r))((h_2\theta_2/(1 + r)) - 1)\theta_1 m_k n/(r + \theta_1 + \theta_2)$

$\partial \Delta w_{2k}/\partial h_1 = 0$

$\partial \Delta w_{2k}/\partial h_2 = \theta_1\theta_2 m_k n/r(r + \theta_1 + \theta_2) > 0$

$\partial \Delta w_{2k}/\partial \theta_1 = ((1 + (1/r))(r + \theta_2)/(r + \theta_1 + \theta_2)^2)((h_2\theta_2/(1 + r)) - 1)m_k n$

$\partial \Delta w_{2k}/\partial \theta_2 = ((1 + (1/r))(1 + (h_2(r + \theta_1)/(1 + r)))\theta_1 m_k n/(r + \theta_1 + \theta_2)^2 > 0$

Note: The inequalities shown presume that θ_2, m_k, $n > 0$.

recur on an intermittent basis. It is consistent with the view that the firms' procedures are not simply devised at the outset, but are improvised when circumstances change and subsequently kept alive as one of the traditions of the firm.

6. Information costs and transaction costs

So far it has been assumed that the information collected has been completely accurate. In general this will not be the case. There are two main reasons for this: incompetence and dishonesty. Incompetence is a problem because what is generally observed when carrying out an investigation is not the thing which is of interest itself, but rather some symptom of it. Symptoms need to be diagnosed correctly if the appropriate inference is to be drawn. Not everyone is equally competent at diagnosis, however, as explained in the following section. This problem is independent of whether the decision-maker carries out his own investigation or not.

Dishonesty, however, is a problem only when information is obtained at second hand. For an autocratic firm, therefore, dishonesty is not a major issue because the owner collects all the key information himself. But for a consultative firm, in which the provision of information from different sources is delegated to different people, it can be a serious problem. The motivations of the individual delegates may not be totally aligned with those of the owner of the firm (Milgrom and Roberts, 1992). If they can anticipate how the information that they supply is going to be used then they may supply false information in order to generate a result which advances their own interests rather than the interest of the owner of the firm.

Dishonesty is a major preoccupation of the theory of transaction costs. Indeed, it could be said that transaction costs theorists have pursued the theme of dishonesty so single-mindedly that they have ignored many other relevant issues in the theory of the firm. The same goes for agency theory – which for the purposes of this paper may be subsumed under transaction cost theory as a special case. As noted earlier, transaction cost theory focuses on the quality of information, and how this can be assured when information is obtained from other people. The quantity of information – in particular, how much flows, and through what channels – has not been properly considered (see Radner, 1992). It is therefore quite important to explore this "grey area" of the interface between information costs and transaction costs in more detail.

Managerial dishonesty can damage the firm in two main ways. The first is that misinformation leads to the wrong decision, so that resources are wasted. The second is that although the correct decision is made, some of the profits that would have accrued to the owner are redistributed to the managers instead.

The main vehicle through which redistribution operates is the budgeting system. Consider once again the problem of choosing the production location in the steady state. The manager at the first location may pretend that conditions there are always bad. This produces misinformation whenever the conditions are in fact good. If the manager at the second location correctly reports that conditions there are good, then production will be diverted to the second location. This misallocation of resource incurs a cost to the firm of $c_{21} - c_{11}$. If, on the other hand, the manager at the second location correctly reports that conditions there are bad then production will remain at the first location but the manager there will receive a budget c_{12} instead of c_{11} which he really requires. If the manager can appropriate this surplus through, for example, expenditure on perks, then the costs of the firm increase by $c_{12} - c_{11}$. Thus dishonesty by the manager at the first location imposes on his employer an expected cost $p_1(p_2(c_{21} - c_{11}) + (1 - p_2)(c_{12} - c_{11}))$. The second component in the main brackets represents the expected gain to the dishonest manager; the first component is a deadweight loss to the firm from the distortion of the output plan.

It is possible, of course, that the manager at the second location will be dishonest as well. Because the second location is not so advantageous as the first, however, the scope for profiting by dishonesty is much reduced. For if the second manager always reports bad conditions then he will never be allocated any production and hence he cannot extract any surplus directly. He may, however, recognise that by reporting bad conditions he assists the first manager to appropriate rents. The two dishonest managers may therefore make a deal

that they will both report bad conditions, and that the second manager will receive a pay-off from the first. Under these conditions production is always allocated to the first plant, even when it should go to the second one, and the production budget is always c_{21}. The manager at the first location generates an expected surplus $p_1(c_{12} - c_{11})$. Some proportion of this – up to, but not exceeding p_2 – is passed on to the second manager. In addition to this loss from inflated budgets, the firm incurs a deadweight loss $(1 - p_1)p_2(c_{12} - c_{21})$ on account of the misallocation of production.

As a firm becomes larger, and more managers are employed, such problems are likely to multiply. The size to which the firm can grow will be determined by how well these problems can be controlled.

One obvious approach is for the owner to collect the information on local conditions himself. If he cannot trust other people then there is little point in asking them for information. The owner therefore becomes an autocrat, collecting all the relevant information himself. This is the converse of the point made earlier: because autocrats do not need to trust other people, those who cannot trust other people will become autocrats themselves.

But what kind of autocrat will the owner become? The costs of investigating conditions himself are likely to be considerable. Quite apart from the travel costs of visiting each site, it is more difficult to become familiar with several sites rather than with just one. This suggests that the owner will economise on investigation costs by switching to one of the less consultative procedures. He will "consult" less with himself than he would with other people (if only he could trust them). He is pushed in the direction of strategy 1, in which he does not investigate conditions at all, but acts on subjective judgement alone. This is particularly frustrating for any local manager that does happen to be honest, since advice he gives the owner in good faith will be ignored.

One reason why local information is so much easier for local managers to collect is that information is often a byproduct of other activities. In particular, information on local conditions in the current period may be a byproduct of the implementation of production plans in the previous period. Unless the owner is going to assume sole responsibility for implementation as well as investigation, it would therefore be better for him to concentrate on improving the honesty of local managers instead, even if significant expense is involved.

Given that the problems of dishonesty are particularly acute when collusion occurs the owner can begin by playing off one manager against the other. He may, for instance, persuade the manager at the second location that a manager at the first location who is dishonest in reporting information will also be dishonest in cheating on any deal between them. A more constructive approach would be to persuade the manager at the first location that the manager at the second location is honest and that conditions at the second location are invariably good. If the manager at the first location believes this then, subjectively, he faces continual competition from a source of supply costing c_{21}. Since $c_{21} < c_{12}$ this means that any false report of bad conditions at the first location will be self-defeating, since the owner will automatically switch production elsewhere.

Of course, the manager could attempt to call the owner's bluff by making a false report and seeing what happens. To counter this the owner could decide that he will always respond to a bad report from the first location by switching production to the second location whether conditions there really are good or bad. This has the advantage that since production is switched there anyway it is unnecessary to investigate what the conditions actually are. This is, in fact, just consultative strategy 2 described above. The owner has switched from strategy 4 to strategy 2 to incentivise the manager at the first location to make an honest report. The integrity of the manager at the second location does not matter because he is not being consulted anyway. There is still a cost, however. For although this strategy eliminates the incentive to cheat, it causes strategy 2 to be substituted for strategy 4 under conditions where strategy 4 would otherwise be preferred.

The most direct way of all to tackle the problem of dishonesty is to eliminate the distortion of incentives built into the budgeting system at the outset – namely that the owner incurs the production costs and that the managers who provide the information on which the decision is based do not. The owner could switch to subcontracting

production to the local managers for a uniform price. This means that the local managers would bear the full financial consequences of production decisions.

In the stable state prices would be negotiated under competitive conditions (though not perfectly competitive conditions) for the reasons explained above. Each manager's information on local conditions would be encoded in his decision on the price at which to tender or, alternatively, in his decision whether to accept work from the owner at some given price. The principle behind this subcontracting arrangement is that the price at which a production contract is agreed embodies a more honest report of local conditions than does the report that would be given by a managerial employee within a conventional budgetting system.

This principle only works well, however, when different managers control activities which are substitutes for each other. If the activities are complementary instead then it does not work so well. Thus competitive subcontracting will not be effective in a volatile state because when demand is high the owner needs to call upon activities in both locations. If each manager knows that the product sells for a price in excess of the cost of production under bad local conditions then with buoyant demand the rents accruing to managers under subcontracting may be even greater than under a conventional budgetary system. This shows that the use of subcontracting by large firms is very sensitive to the degree of substitution in production. When complementarity rather than substitution prevails, therefore, an autocratic approach may prove unavoidable even in the stable state in order to control transaction costs.

7. Entrepreneurial judgement, information cost and hierarchical structure

It was noted in the previous section that incompetence is a major constraint on the quality of information, and that this applies whether information is obtained from other people or direct from its source. The question of competence is related to the form in which information appears. If the subject matter being studied generates a clear and unambiguous signal of the state it is in then it normally takes little skill to observe it and record it. Speed, accuracy, and a good command of language may well help, of course. But the situation is very different from where only a symptom, or a collection of symptoms, is available, and diagnosis is required before any inference can be drawn. In this case considerable background knowledge may be required to interpret the symptom. Observation and recording skills are no longer the most important skills involved.

It seems to be characteristic of volatile environments that information on their current state often comes in this symptomatic form. Once the symptoms have been interpreted, moreover, it may be possible to express the result only in rather general terms.

Conversely, it is a characteristic of stable environments that fluctuations in them are often signalled in an explicit way. Fine-grained distinctions between different situations can be drawn and it is often quite appropriate to express the results in numerical form.

It is sometimes more appropriate to think of the interpretation of symptoms as reducing uncertainty rather than as providing definite information – although from the standpoint of subjective probability these are really just two sides of the same coin. This perspective correctly emphasises, however, that the communication of the interpretation needs to be accompanied by probability statements which indicate how much confidence can be placed in the report. In practice, probabilistic information is often tacit, in the sense that it is difficult to explain to other people. This significantly increases the cost of communication and encourages the owner of the firm to collect as much information of this kind as possible for himself.

The key factor in a volatile situation was identified in the previous model as the state of demand. This identification seems plausible in the light of the preceding remarks, in that the state of demand often reveals itself through symptoms – in particular, qualitative symptoms of fashion trends and social change. Although demand can, of course, be estimated at the end of a period from actual sales, it is demand at the start of a period that is relevant here. The idea that the owner takes responsibility himself for interpreting the symptoms of demand accords with the view that

this interpretation would prove difficult to communicate if it were made by anyone else.

In the stable state the owner continues to monitor demand, but he now synthesises demand information with information on production conditions, because demand information is no longer decisive for the output plan. Production conditions generate more explicit symptoms than demand conditions because of the relative objectivity of technology and local resources compared to customers' tastes. The assumption in the model that the owner delegates the investigation of production decisions to local managers accords with the idea expressed above that explicit information is easier to communicate than symptomatic information.

Not only is the communication of production information easier, but its collection requires rather different skills as well. Whilst local managers may lack the owner's ability to interpret symptoms of demand, they may have an expertise he does not possess in reporting productions with technical accuracy. This suggests that in the consultative firm there will be some degree of specialisation within the management team according to the type of information that needs to be provided. It is advantageous if the person ultimately responsible for the synthesis of the information also provides the information that is most costly to communicate – which normally means the information on the state of demand. Because this information on demand is normally so vital – it is, indeed, decisive in the volatile state – it is most important that it be correct. This means that the person who provides it must be well incentivised to tell the truth – a consideration that is even more important where this person acts as synthesiser too. The person who is best incentivised in material terms is, of course, the owner himself. The owner will therefore take responsibility both for the provision of information on demand and for the synthesis of this information with the production information that his local managers provide.

This arrangement generates a simple hierarchical structure for the consultative management team. The local managers report information up to the owner, who synthesises this information with his own information on demand to determine the output plan. The output decision regarding each location is then communicated back to the local manager, who is responsible for its implementation. As a byproduct of this implementation, information relevant to next period's production conditions is obtained. This forms the basis for next period's report to the owner.

This consultative hierarchy may be contrasted with the autocratic hierarchy in which information on production conditions is ignored by the owner on the grounds that it is either too costly to collect or, even if available, would not affect the optimal decision. In the autocratic hierarchy information required for implementation is still fed down from the top: each local manager is told the output plan for his own location. But information is not fed up the hierarchy at the same time. Autocratic hierarchies involve a one-way flow of information, therefore, in contrast to consultative hierarchies where the flow of information is two-way.

The location of entrepreneurship in the hierarchy is rather different too. In the autocratic hierarchy all entrepreneurial activity is clearly concentrated at the top. In the consultative hierarchy some entrepreneurial activity can be delegated to lower-level members of the management team. How much is delegated really depends on how much judgement is required in interpreting the symptoms on which production information is based. If production information is quite explicit – as assumed above – then lower-level management is a technical activity rather than an entrepreneurial one. But if it is symptomatic – like demand information – then considerable judgement may be required. In this sense, the consultative firm then becomes a team of entrepreneurs (Wu, 1988).

One of the consequences of increasing international integration in the post-war period has been that firms have acquired the ability to produce in varied locations. These locations differ not only in tangible and measurable ways – for example local wages – but in intangible and qualitative ways as well – such as local culture. This has increased the entrepreneurial input to reporting local production conditions, and so has shifted large firms away from hierarchical structures based on technical management towards hierarchies based on delegated entrepreneurship instead. Similar remarks could be made about the growing dispersion of marketing information sources within the global-

ising firm, which has further increased the dispersion of symptomatic information.

Small business research needs to take account of the fact that because of these changes many of the large firms with which small businesses are conventionally compared have transformed their organisational strategies during the last two decades. Their consultative procedures now synthesise information which is supplied in the form of entrepreneurial judgements reported by managers at widely-dispersed locations. Their structures have become less bureaucratic and more collegial as a result. They have acquired some of the virtues that business folklore associates with small firms instead.

8. Leadership and information cost

One of the presumed advantages of the small firm is the enhanced opportunity for face-to-face communication. There are two aspects of this that need to be distinguished. The first is the ability to improve the implementation of a given decision. This applies whether or not the firm is autocratic. The second is the ability to improve the flow of information from managers to the owner of the firm. This is relevant only to the consultative firm. It follows immediately from this that an autocratic style of management fails to exploit fully one of the key advantages of being small, namely the ability to improve consultative procedures.

This suggests that there are, in fact, two quite distinctive reasons for being small. One is because autocratic management is efficient, and the other is because consultation depends crucially on face-to-face communication. This is borne out by the importance of professional partnerships in the small firm sector, and their distinctive style of organisation. While in the manufacturing sector small often means autocratic, in professions such as law, accountancy and medical practice, small often means consultative because of the importance of face-to-face communication amongst expert colleagues.

The impact of face-to-face communication is rather different when consultation rather than implementation is involved. In implementation the question is whether orders have been properly understood and honestly carried out. Understanding orders is not usually much of a problem because orders can usually be made quite explicit. For example, it is not difficult for the owner to explain to a local manager what production level is required.

Checking up on honesty in implementation is slightly more difficult, though the issue is still a fairly straightforward one. It is usually a question of ensuring that the requisite effort has been applied. The problem is typically that a lazy manager will not have carried out the task properly, and that he will try to disguise the fact by withholding information about what he has really done. Face-to-face communication allows the owner to supervise the implementation process, and so prevent the concealment that can occur when the manager operates at a remote location. This is the conventional view of how supervision overcomes the "agency problem".

Where consultation is involved, however, other considerations apply. For a start, the information that flows as part of a consultation process may be of a symptomatic rather than explicit kind. As noted above, this is most likely to apply when production locations are widely dispersed. Thus the geographical dispersion of production actually increases the demand for face-to-face communication – a paradox which can only be resolved (in the absence of teleconferencing) by managers travelling frequently to special meetings. The important point to note, however, is that the more dispersed the entrepreneurial function becomes within the hierarchy, the greater is the need for face-to-face communication.

The second point is that where consultation is concerned, dishonesty does not involve withholding information so much as supplying misinformation instead. Moreover, misinformation may be more difficult to tackle at source because it is not the result of lack of effort so much as strategic calculation. Thus there is no simple manifestation, such as shirking, which a supervisor can monitor. The only real check is to replicate the information itself.

Replication is very expensive for the owner, and there is always the risk that his competence may not be as great as the manager's, so that the manager may be charged with dishonesty simply because the owner himself has made a mistake. If the owner delegates the replication to another manager then he simply has two different accounts

of the same situation to compare and no objective way of deciding, in the event of a disagreement, which of them is right.

Given the limitations of supervision, it would be nice if the manager could be persuaded to supervise himself. This highlights another distinctive aspect of face-to-face communication which has so far been overlooked in the discussion. This is the capacity for face-to-face communication to develop emotional bonds, which encourage the manager to be honest with the owner out of loyalty to him. The manager knows he will feel guilt if he cheats the owner, and the threat of this self-inflicted emotional penalty encourages him to be honest instead (Casson, 1991).

This emotional mechanism helps to ensure integrity in both consultation and implementation, but its role is more crucial in consultation than in implementation because the alternative of supervision by the owner is less effective in this case. This reinforces the point made above that face-to-face communication is most important in the consultative firm. In the autocratic firm the problem of dishonesty arises only in implementation, and dishonesty in implementation can often be controlled through supervision. Hence the autocratic firm does not need emotional bonds to the same extent as the consultative firm and therefore does not have the same need to exploit face-to-face communication either.

The link between face-to-face communication and consultation that is forged by emotional bonding is further reinforced by the link between consultation and a sense of loyalty to the firm. When managers are consulted their self-esteem is raised, because consultation signals that they are believed to be competent and are trusted to supply honest information (see Section 6). This gives them a sense of genuine participation in decisions. Face-to-face communication, it can be argued, heightens the emotional rewards from participation – provided, of course, that consultation takes place within a friendly atmosphere. Recognising the emotional benefits they derive from participation, managers desire to reciprocate. The most obvious way they can do this is to offer their integrity.

Note, furthermore, that because consultation strengthens integrity, it spills over to implementation too. Managers who offer integrity to the firm will not only supply honest information but will work hard to implement output plans too. Thus through the mechanism of the emotional bond, face-to-face consultation indirectly improves implementation. Firms that find the supervision of implementation difficult may therefore choose to consult employees face-to-face, not because they need the information, but simply because they need to guarantee effective implementation instead. Thus even where a volatile environment might favour an autocratic management style, a face-to-face consultative approach may be employed in order to assure the quality of implementation.

9. Summary and conclusions

This paper has embedded the analysis of small business behaviour within a general theory of the organisation of the firm. It has been noted that small firms often manifest an autocratic management style and that this may be interpreted as a rational response to an environment in which there is a single major source of volatility. An autocratic management style may also reflect an inability to trust the information supplied by other people, or a lack of intellectual ability to devise and remember more sophisticated procedures of consultation. Autocratic management is favoured when implementation is easy to supervise, for then employees do not need to be consulted just in order to strengthen their loyalty to the firm.

Theory suggests that small firms in professional services are very different from small manufacturing firms in that they have particular reasons for adopting a consultative management style. Their managers need to synthesise symptomatic information from different specialised sources, and this requires a collegial approach to decision-making.

Large firms tend to adopt a consultative management style because they face many different sources of volatility rather than a single dominant one. When their activities are very dispersed then information on local conditions is likely to be symptomatic too, and this favours face-to-face communication. Rather ironically, therefore, highly-dispersed large firms will strive to achieve a "small firm" feel amongst their management

team. By a further irony, this "small firm" feel is not, in fact, the kind of "feel" that is actually experienced within many small firms because of their autocratic style.

Consultative styles are embodied in procedures which are normally established by precedent. These procedures are geared to dealing with relatively stable conditions, and may be over-ridden from time to time by autocratic control during spells of volatility. It is important that they are remembered, however, because otherwise they must be re-invented each time stable conditions recur.

All firms are prone to intermittent spells of volatility. For small firms volatility is the norm whereas for large firms it is rather the exception. Because consultative procedures are more difficult to devise and to remember than autocratic ones, small firms are likely to persist with autocratic procedures through brief spells of stability, whereas large firms are more likely to switch procedures whenever circumstances change. This suggests that although small firms are relatively flexible in their ability to cope with a single major source of volatility, this flexibility does not extend to the procedures of the firm itself. Large firms are more flexible in this latter respect because they can always over-ride normal procedures for a time and then return to them later.

References

Casson, Mark C., 1991, *Economics of Business Culture*, Oxford: Clarendon Press.

Casson, Mark C., 1994a, 'Information Costs and the Organisational Structure of the Multinational Enterprise', Paper presented at the Annual Meeting of the Academy of International Business, Boston, November 1994.

Casson, Mark C., 1994b, 'Why are Firms Hierarchical?' *Journal of the Economics of Business* **1**, 47–76.

Coase, Ronald H., 1937, 'The Nature of the Firm', *Economica* (New Series), 386–405.

Milgrom, Paul R. and John Roberts, 1992, *Economics of Organisation and Management*, Englewood Cliffs, N.J.: Prentice-Hall.

Radner, Roy, 1992, 'Hierarchy: The Economics of Managing', *Journal of Economic Literature* **30**, 1382–1415.

Williamson, Oliver E., 1985, *Economic Institutions of Capitalism*, New York: Free Press.

Wu, H., 1988, *Production, Entrepreneurship and Profits*, Oxford: Blackwell.