CONTRACTS, COOPERATION AND TRUST:
THE ROLE OF THE INSTITUTIONAL FRAMEWORK

ESRC Centre for Business Research, University of Cambridge
Working Paper No. 10

Simon Deakin
Faculty of Law, and
ESRC Centre for Business Research
University of Cambridge
Sidgwick Avenue
Cambridge CB3 9DE

Tel. 01223 335244
Fax. 01223 335768
E-Mail sfd20@econ.cam.ac.uk

Frank Wilkinson
Department of Applied Economics, and
ESRC Centre for Business Research
University of Cambridge
Sidgwick Avenue
Cambridge CB3 9DE

Tel. 01223 335244
Fax. 01223 335768
E-Mail sfw11@econ.cam.ac.uk

September 1995

This Working Paper relates to the CBR Research Programme on Corporate Governance, Executive Pay and Economic Performance.
Abstract

This paper examines the nature of contractual cooperation within and between firms in vertical supply chains. We suggest that such cooperation should be conceptualised not as a consequence of but as an alternative to market-based competition between economic agents. The quality of cooperation between agents is then dependent upon the presence of norms, conventions and other institutional arrangements which foster ‘trust’, particularly by controlling the process of competition through product and labour-market standards. One implication is that arrangements which are commonly condemned as collusive or restrictive may, for this reason, have important efficiency-enhancing effects.

Acknowledgements

The support of the Economic and Social Research Council is gratefully acknowledged. The work was part of the ESRC Contracts and Competition Programme and was funded by ESRC award number L114251016. We are grateful for the advice and comments of our colleagues Reinhard Bachmann, Tom Goodwin and Christel Lane.
CONTRACTS, COOPERATION AND TRUST: THE ROLE OF THE INSTITUTIONAL FRAMEWORK

1. Introduction

The new institutional economics has posed the question of why certain transactions are internalised within firms and others remain in the realm of the market, without as yet producing a general theoretical framework which can satisfactorily explain the diversity of different contractual forms and their welfare implications. The idea of a sharp firm-market dichotomy has had to be modified by studies of network relations which are ‘neither market nor hierarchy’ (Powell, 1990). The concept of a network or ‘hybrid’ form of industrial organisation is particularly applicable to ‘relational’ contracts or linkages between firms, which may be classified as either vertical (such as franchising or customised subcontracting) or horizontal (such as joint ventures between firms and looser forms of cooperation based on membership of trade associations or ‘industrial districts’). Relational linkages of this kind, which are characterised by a high degree of mutual inter-dependence between autonomous contracting agents, conform neither to the model of exchange under conditions of atomistic competition, nor to that of the hierarchical ‘authority relation’ of the vertically integrated firm.

This paper seeks to throw light on the comparative properties of these different forms of economic organisation by considering them from the point of view of cooperation in productive systems. An institutional economic framework is developed to explain how cooperation between economic agents is an essential element in the *dynamic efficiency* of a productive system, measured by its capacity to enhance information flows and engender a high rate of innovation in response to external change. It will be suggested that relational contracts, displacing decentralised exchange, are an important but in themselves incomplete means of promoting cooperation of the kind needed to ensure dynamic efficiency. The degree to which such contracts are self-enforcing is limited; nor can market competition be expected to provide a viable alternative in circumstances of high transaction costs. This may imply a certain role for
legal and extra-legal (or ‘social’) norms, but the strict enforcement of promissory obligations is only one aspect of this and, in successful productive systems, among the least important. A growing body of evidence suggests that greater significance should be accorded to norms and conventions whose functions include the regulation of market power and the placing of limits on ‘destructive competition’. Norms of this kind, which mainstream economic thought tends to see as the product of ‘collusion’ or rent-seeking, may instead be important means of fostering ‘trust’ between contracting agents and hence enhancing overall economic welfare.

2. Technical Interdependency and Cooperation in Productive Systems

Economic thought places strong emphasis on specialisation and individualism. Specialisation by way of the division of labour is seen as the central driving force of economic progress and individuals are seen as motivated by their own self interest to engage in economic activity. Yet even a cursory glance at the production process reveals a high degree of technical interdependence between workers and between labour and the means of production (including sources of financing), and a more careful consideration suggests that the more detailed the division of labour and the more highly specialised the factors of production, the greater becomes the need for those factors to work closely together.

It is useful to think of productive systems as consisting of the stages of production of particular products and the linkages between the different stages (Wilkinson, 1983; Tarling and Wilkinson, 1987). Two forms of production linkages can be identified: those within the vertical supply chain and those extending out horizontally from particular points in the chain (Wilkinson and You, 1992; Deakin, Lane and Wilkinson, 1994). *Vertical linkages* join together the stages in the process of production, which typically include the preparation of raw materials, the manufacture of semi-finished products, the finishing and assembling of products and the packaging and forwarding of finished goods towards the final market. *Horizontal linkages* exist between processes in the vertical chain and products and services which are particular to that stage, e.g. with the
producers of machines and the suppliers of training, technical, administrative and other services. Perhaps the closest analogy is with a river and its tributaries. What we call here horizontal links have their own vertical chains of products and services whose final stages are specific to intermediary stages, but whose productive contribution merges with and adds to the main vertical flow of production of the product or products which define the system in question.

At each stage the different types of labour, machines, materials and services which are combined in the chain of supply are technically interdependent: one cannot operate without the other so that the failure of any adequately to perform its productive function lowers the joint product of the whole. This technical interlocking extends to relations between the different stages of the chain. The requirements of products and processes define to a certain extent the attributes of the labour force and the means of production by which they are to be produced. How well the latter are designed and made (or trained in the case of labour), the timeliness of the delivery of constituent parts and the care by which products are fashioned and assembled will determine the quality of output. The operational efficiency of productive systems - that is to say, the effectiveness of the utilisation of productive resources - will depend on how well labour and the means of production combine within each stage of production and how well the different stages of production work together. Consequently, cooperation between the agents of production (the owners and/or controllers of labour, the means of production and the different stages of production) is a technical necessity determining the quantity and quality of the output of the productive system and hence its competitiveness.

A central feature of the necessary co-operation within productive systems is the sharing of information in response to conditions of pervasive risk and uncertainty. Risk and uncertainty arise from the problems of matching production processes to shifts in demand in intermediate and final product markets, and increase considerably to the extent that products are either customised in the sense of being produced to a particular customer’s specification, or new in the sense of embodying innovative techniques and processes. The avoidance of producing unsaleable stock or allowing
demand to go unsatisfied requires responsive linkages within the chain and the more uncertain final demand is, the more important this will be. Success in production and marketing will also depend on access to information on changes in products, processes and forms of organisation in rival systems. Here cooperation is important because of the problem-solving benefits of working together and because the sharing of information increases the pace of diffusion of new processes and products and hence the pace of technical progress of the productive system. Thus in addition to enhancing operational efficiency, cooperation is essential in order to enhance the *dynamic efficiency* of the system in terms of its capacity to adapt to a changing external environment and maintain or increase its rate of technical change and innovation.

A ‘productive system’ in the sense that we have used that term could be organised on the basis of a single, vertically integrated firm, or alternatively through inter-firm relations of various kinds, both horizontal or vertical. Depending on the extent of vertical integration, there will be different arrangements of property rights and different varieties of contractual relations may emerge. The question then is: what is the contribution of these alternative organisational arrangements to the securing of cooperation in production?

3. **Competition and Self-Enforcement as means of Ensuring Cooperation**

The standard answer given to this question by mainstream economics is that cooperation is secured through market competition. Markets provide the opportunities for individual agents to trade and if there are sufficiently large numbers of agents, each with a large number of opportunities for alternative productive employment and each is equally well informed, the market will operate freely to ensure cooperation and determine the relative prices of the factors of production. Each of the agents will secure the compliance of the others in production because they can choose alternatives from amongst a large number of substitutes. Meanwhile, the price of factors of production will be determined by the strength of competition for their services as set, in turn, by the level of supply, the structure of demand for products and the factor composition of the techniques of production.
This answer must be regarded as unsatisfactory, however, if the essential insight of transaction cost economics - that monopoly, sunk costs and related forms of market failure are pervasive in complex exchange processes - is to be taken seriously. The problem of monopoly extends beyond the fixing of prices in the market into the production process itself, thanks to measurement problems of various kinds which have been analysed in the economics of information. The value of information is only revealed in use, or once the exchange intended to fix its value has taken place (Arrow, 1962). More generally, the failure of contract performance to live up to expectation is traceable to difficulties of observation and measurement both *ex ante* (‘adverse selection’) and *ex post* (‘moral hazard’) (Akerlof, 1970; Arrow, 1985). The involvement of agents in the production process creates or reinforces elements of partial or temporary monopoly. Experience in production may give privileged access to knowledge and/or may create knowledge by learning by doing which may or may not be available to would-be competitors or to partners in production. Moreover, market exchange and the subsequent productive activity may transform a general into a specific asset or, at least, add specific attributes to an asset which has general uses. For example, general purpose machines may be modified for particular uses or workers may receive on-the-job training or acquire by work experience specific skills necessary to operate particular machines, operate within particular organisations or to produce particular goods or services. The expectation that sunk costs arise *ex post* in these ways may deter the parties from contracting *ex ante*, and so, more generally, create a barrier to entry into a particular contract and, by extension, into a particular market (see generally Williamson, 1975, 1985).

One response to the presence of high transaction costs in decentralised markets is vertical integration through the firm: to use the terms chosen by Coase ([1937] 1988), the ‘authority relation’ of employer and employee ‘supersedes’ price as the mechanism for coordinating economic resources, and ‘conscious power’ displaces impersonal, arms-length exchange. One way of expressing this insight is to say that *competition* and *hierarchy* are alternative forms of economic coordination, and that the *market* and the *firm* are the alternative governance structures to which they correspond. The notions of *cooperation* and *networks* can then be added as a third and,
until recently, comparatively neglected set of alternatives (e.g. Loasby, 1994: 49; Powell, 1990). The concept of cooperation has been neglected because both mainstream economic theory and large parts of the transaction cost literature have taken the view that market competition suffices, in the long-run, to ensure the evolution of optimal organisational forms. But if this view is taken, the notion of cooperation is superfluous to the analysis: cooperation is once again, just the outcome of competitive processes, and organisational forms do not independently influence it.

There is an unresolved ambiguity in the approach taken to this question by transaction cost theory and the associated principal-agent models of contracting. It is not clear how far these approaches see the firm as an alternative to the market. If it is maintained that ‘firms are ultimately formed and revised in markets’ as a leading account suggests (Klein, Crawford and Alchian, [1978] 1986: 249), hierarchy fades away as an analytical category. The logical conclusion of this approach is that there is nothing to differentiate the employment relationship from other varieties of contractual relation (Alchian and Demsetz, 1972). In the principal-agent model of the firm, contracts are seen as responses to agency costs of various kinds arising out of asymmetrical information and the need for monitoring; contractual forms will emerge as the product of efficient choices by economic agents. It is only necessary to assume the prior existence of a legal system in order to supply property rights as the subject-matter of exchange. The firm is just a legal fiction which serves as the focal point of a ‘nexus of contracts’ between the owners of different inputs (shareholders, lenders, employees, suppliers) and the firm’s customers: ‘the “behaviour” of the firm is like the behaviour of a market; i.e., the outcome of a complex equilibrium process’ (Jensen and Meckling, 1976: 311). But why should such an equilibrium be expected to emerge on the basis of the efficient choice, by the agents concerned, of various interchangeable contractual forms, given the existence of asset specificity and monitoring and information costs which are, by definition, a significant barrier to the workings of the price mechanism?4

If competition is neutralised, contracts may either be self-enforcing, thanks to reputational or similar effects, or can be enforced through the legal
liability rules of contract law. We would suggest, however, that neither self-enforcement nor legal enforcement provides an adequate, general explanation to the problem of how cooperation is achieved in long-term contracts which are by their nature insulated, to a greater or lesser degree, from competitive forces (Wilkinson and You, 1992; Deakin, Lane and Wilkinson, 1993; 1994).

As far as self-enforcement is concerned, the theory of repeated games suggests that cooperation may be enforced by the threat of retaliatory defection by one of the trading partners against the other. Although the conditions in which the threat of defection will hold are quite restrictive — some form of long-term relationship involving repeat trading and reputational effects is needed — they are more likely to be satisfied in the circumstances being considered here, namely long-term relations between economic agents involving the creation of a transaction-specific surplus from exchange and the putting in place of well-functioning mechanisms for the dissemination of information. Where there are transaction specific rents of this kind, contract terms can be used to create a material incentive structure designed to punish any breach through such devices as collateral, ‘hostages’ or ‘hands-tying’ (Williamson, 1983; Kronman, 1985). Equally, the contract may contain an incentive structure incorporating a positive reward system of some kind in order to elicit cooperation over the long-term: employment contracts containing job security guarantees and pension entitlements in return for long service are an example of this, as analysed by efficiency wage theory (Bulow and Summers, 1986). However, it is not necessarily evident that contracts of this kind can rest solely on self-enforcement. The problem is that any such contract may be subject to renegotiation in mid-performance to the benefit of the party threatening defection or breach, thereby undermining the credibility of the self-enforcement strategy which it embodies (Harsanyi and Selten, 1988).

Alternatively, defection may be deterred and performance encouraged by the emergence of a moral or social norm beyond the contract itself, lending credibility to a threat of retaliation or representing an independent source of motivation (Axelrod, 1984). What is not clear in this approach, however, is why such social norms emerge in certain social and industrial contexts.
and not in others, despite there being a general recognition, in the latter, of the need for cooperation between economic agents in order to improve general welfare (Lorenz, 1988; Wilkinson and You, 1992). It seems doubtful that the answer to the question ‘why are there no industrial districts in Britain’, for example (Zeitlin, 1994), can be answered without a deeper examination of the historical and institutional forces at play, including industrial policy, than game-theoretical accounts can supply.

In principle, legal enforcement of contractual obligations provides a means of underpinning the parties’ mutual expectations and granting additional security to their agreement, without the state encroaching on their private rights. The courts’ role does not purport to go beyond identifying and enforcing the contract the parties themselves have made, so that contract law can be seen respecting the autonomy of the parties by providing a space for private ordering (Epstein, 1975). However, theoretical and empirical studies alike suggest that the liability rules of contract or commercial law provide only a partial solution, at best, to the problem of non-cooperation in contractual relations (Macaulay, 1963; Beale and Dugdale, 1975; Charny, 1989; Bernstein, 1992; Campbell and Harris, 1992; see generally Deakin, Lane and Wilkinson, 1994: 337-340). Information and measurement problems make it difficult for the courts effectively to gauge the value of the parties’ ex ante contractual expectations; nor can they enforce or supervise specific performance of a contract without encountering serious problems of moral hazard. As a consequence, ‘the legal right to enforce a promise can reduce but not eliminate the insecurity associated with all temporally asymmetrical exchanges’ (Kronman, 1985: 25). Strict legal enforcement may, indeed, be actively harmful in the context of long-term, relational contracts which rest upon shared assumptions and understandings and on the possibility of extra-legal sanctions which may not be expressed in formal terms. Such arrangements may easily be misinterpreted by the courts, which to not have access to the specialised knowledge or assumptions shared by the parties. In short, the availability of legal enforcement can only provide part of the answer to the question of how cooperation is enforced in long-term contractual relations. It is necessary to take a broader perspective, one which identifies the distinctive properties of different forms of inter-and intra-firm contracting and which
explains the role played by the institutions which regulate these contractual forms.

4. Cooperation in Intra- and Inter-Firm Relations

The debate in economics about the nature of the firm, which of course long pre-dates Coase, mirrors the wider debate over the institutional origins of cooperation. A view which goes back at least to Adam Smith is that self-interest on the part of economic agents, coupled with the division of labour and a legal system which respects the property and contract rights of individuals, is sufficient to achieve the necessary cooperation in production. As individuals become increasingly specialised as producers, they become more dependent on others as consumers. Thus, they are ‘at all times in need of the co-operation and assistance of the great multitudes’ (Smith, 1974: 118). But these needs are met, Smith argued, by exchange motivated ‘not by the benevolence of the butcher, the brewer or the baker... but from their regard for their own interest’ (ibid: 119). The pursuit of individual self interest in exchange in a free market thereby secures the level of cooperation necessary for society as a whole.

An emphasis of the importance of cooperation in production came from an early criticism of Adam Smith’s treatment of the division of labour. Edward Gibbon Wakefield (1835) in his editorial notes to The Wealth of Nations argued that Adam Smith confused the categories of labour and employment, the latter meaning the work performed by labour. Labour, Wakefield argued, is naturally divided into pairs of hands so that the division of employment between many pairs of hands requires workers to work collectively: ‘The greatest division of labour takes place amongst those exceedingly barbarous savages who never help each other, who work separately from each other; and the division of employment with all its great results, depends on combinations of labour and cooperation... all improvements in the productive power of labour, including the division of employment depend upon cooperation’ (Wakefield, 1835: xx). Wakefield went on: ‘Cooperation appears to be of two distinct kinds, such cooperation as takes place when several persons help each other in the same employment; secondly, such cooperation as takes place when several persons help each
other with different employments. These may be termed simple cooperation and complex cooperation.’ Cooperation is to be found in the workshop: ‘the division of employment which takes in a pin factory results from and is wholly dependent on, the union, generally under one roof, of all the labour by which pins are made’ and more generally is ‘dependent also upon arrangements, agreements, concert or combination of a general kind, in which the whole of society takes part’ (ibid.:29). For Wakefield, then, growing cooperation both between producers and consumers and between producers was dependent on social arrangements between economic agents, including the ‘union’ encapsulated in the firm.

Marx echoed Wakefield’s distinction between simple and complex cooperation and the idea that cooperation existed both within the workplace and at the wider societal level. Marx followed Smith in emphasising the role of the market in ensuring societal cooperation but took the notion of cooperation in production as the basis for his analysis of the labour process. Capitalist employers assembled workers under their command, coordinated their cooperative activities and exploited them by extracting from them surplus value. By bringing workers together in simple cooperation they increased productivity without making any further contribution.7 Assembling labour together provided the opportunity for the division of labour and the evolution of new productive techniques in the manufacturing stage of capitalist development (Marx, 1974, Chapter XIV). Up to this point workers remained capable of an independent economic existence although they became more productive if bought into cooperation; but this changed as machinery and modern industry developed:

In simple cooperation, and even in that founded on division of labour, the suppression of the isolated, by the collective, workman still appears to be more or less accidental. Machinery... operates only by means of associated labour, or labour in common. Hence the cooperative character of the labour process is, in the later case, a technical necessity dictated by the instrument of labour itself (p. 364.)

Within marxian analysis, incentives for labour to participate in production are provided by the securing of the means of subsistence (which are also
owned by capital) whilst the incentive for capital is the extraction of surplus value which arises because the productive power of labour (its use value) is greater that the social cost of reproducing labour power (its exchange value). Capital generates surplus value by controlling the labour process. In the organisation of production capital had dual roles. The first of these is the management of the technical relations of production: how to ensure the highest level of saleable products by the most efficient use of labour and the means of production at its disposal. The ‘technical’ managerial function of capitalism is a necessary but not sufficient condition for generating surplus value. The second element in capitalist organisation can be described as its ‘social’ managerial function by which surplus value is generated by inducing a higher level of labour input than that needed for subsistence and paid for by wages. Through these means capital secures for itself the surplus from cooperative productive activity between workers.

Marshall similarly acknowledged the importance of cooperation in production in the sense that Wakefield and Marx used it8 but did not stay ‘to consider all the implications of this argument’ (1920: 72). However, he placed emphasis on the need for more sophisticated forms of coordination as the division of labour led to increasingly differentiated labour and machines (Marshall, 1947: Book IV, Ch VIII), and paid close attention to the role of organisation in the coordination of the increasingly specialised and mutually dependent productive activities.9. Thus whilst Marshall saw freedom of industry and enterprise10 as a central motivating and integrating force, market success depended on increased specialisation and the development of more effective industrial organisation, a process driven by the entrepreneurial and innovating owner manager.

Marx and Marshall are both part of the tradition in economics in which the hierarchically organised firm plays a central role in not only organising production but also in dictating the pace and direction of economic development. Leading members of this tradition also include Knight for whom risk bearing justified command; Schumpeter for whom innovating entrepreneurs secured ascendancy, if only temporary, by innovation; and Penrose whose firms reaped increasing returns from expanding managerial teams. Here entrepreneurs and their organisations play a central role both
in organising production and in developing markets, and, although the market and the firm may be alternative ways of organising production they are also complementary in that process. From the perspective of this tradition, Coase’s question - ‘having regard to the fact that if production is regulated by price movements production would be carried out without any organisation at all, well might we ask, why is there any organisation?’ (Coase 1988a: 36) - is, strictly speaking, meaningless; Coase’s question takes on its meaning from his initial position that ‘an economist thinks of the economic system as being coordinated by the price mechanism’ (1988: 34).

If a neoclassical, general-equilibrium approach is adopted, then as we saw above it is logical to question Coase’s insistence on the distinctive nature of the contract of employment in terms of an ‘authority relation’. Property rights theory has attempted to reconcile Coase’s initial insight with the criticism levelled against it by Alchian and Demsetz (1972) by focusing on the employer’s ownership rights in the non-human assets of the firm as the central issue from the point of view of incentives (Hart and Moore, 1990). The employer gains compliance with his instructions, it is said, not by virtue of the contractual form of the employment relationship, but through the firm’s ownership of the technical and material assets of production; this differentiates the employer from one who is simply a customer. Although property rights analysis ‘follows Alchian and Demsetz in not distinguishing between the contractual form or nature of sanctions in the two relationships’, it nevertheless uses the notion of ownership as a right of exclusion to ‘capture the idea that one agent is more likely to do what another agent wants if they are in an employment relationship than if they are independent contractors’ (Hart and Moore, 1990: 1150).

But it may be that property rights theory explains too much here: it is not just the employees of a firm who may be excluded from ownership of the physical assets which are involved in a process of production; the same is true of many independent contractual relationships involving the retention of property rights by the buyer. In business format franchising, for example, the franchisor will frequently retain rights not simply in the brandname of the product but also in various intellectual property assets connected with
business know-how. These will be protected by contractual non-
competition clauses which the courts normally uphold. It is also common
for the franchisor to take control of the physical assets himself, by for
example taking a head-lease of the business premises from which the
franchisee works (Felstead, 1991). In customised sub-contracting it is rarer
but by no means uncommon for the customer to retain property rights in
the machinery, materials and designs used (Semlinger, 1991).

On the other hand, the property rights approach does not do enough to
explain the distinctive features of intra-firm relations. While the vesting
of the property rights in the means of production in the firm is clearly one
important source of the asymmetrical nature of the employment relationship,
is it the only one? Ownership rights give the employer the power to exclude
the employee from the production process, whether temporarily through
disciplinary sanctions or more completely through dismissal. But the
existence of power by one contracting party, in the form of a threat carrying
negative connotations for the other, has important limitations as a means
of eliciting cooperation. High monitoring costs must be borne. These can,
however, be averted or lessened if power is depersonalised, in the sense of
being embodied in a system of rules or norms which mediates its exercise
and which introduces a degree of predictability and stability into the relation,
making possible the development of mutual trust. We would suggest that
it is the presence of norms of this kind, rather than the pre-contract
distribution of property rights between agents, which most clearly separate
the employment relationship from other forms of contractual relation.

In this regard, contractual form is important in two senses: the employment
contract underpins the employer’s power of direction, the ‘managerial
prerogative’, while at the same time granting the employee certain
expectations of continuing employment and income security. The principal
distinctive feature of the contract of employment is that the employee cedes
to the employer a degree of authority over the pace and organisation of the
work which he or she undertakes to do. It is not simply the distribution of
property rights pre-contract, but the contract itself which is asymmetrical
and which engenders economic power. The employer does not need to
gain the consent of the employee to any and every change in the way the
job is carried out. While the employer’s inherent powers are subject to limits, in the sense that there are many aspects of the individual contract which are governed by either individual or collective bargaining, employment law recognises that the employer has certain residual rights to the ‘cooperation’ of the employee in areas not covered by explicit or implicit agreement. The residual power which vests in the employer can be seen as a layer of status ‘beyond contract’ (Fox, 1974) which attaches to the relationship both as a matter of law and of social convention. To speak of employment as a continuous process of ‘renegotiation’ or renewal of contracts between parties to a continuing relationship (Alchian and Demsetz, 1972) is to ignore this essential feature: the employer’s implicit rights make regular renegotiation unnecessary.

The open-ended character of the employee’s commitment can be seen as providing the basis for the flexibility needed to ensure co-operation between a large number of agents in the productive process; in Coasean terms this accounts for the firm’s efficiency properties as against those of the decentralised market, where continuous re-negotiation between large number of agents would indeed be necessary. The same idea, in a different form, is found in the marxian distinction between labour and labour power, which has quite different implications: it is because the worker contracts to submit himself or herself to the employer’s orders for the duration of the contract term that the resulting performance of the contract gives rise to the possibility of expropriation of the surplus from production. However, the persistence of the employment relationship notwithstanding this essential asymmetry, and potential instability, is arguably attributable to norms which create a second layer of status to add to that provided by the employer’s implicit powers of direction, namely norms of employment and income protection deriving from labour legislation and collective bargaining between trade unions and employers (Mückenberg and Deakin, 1989; Streeck, 1992). Some transaction cost analyses also recognise that it may be employers’ interest to introduce many of these benefits in order to elicit long-term cooperation (Williamson, Wachter and Harris, 1975).

In these ways the fundamental inequality of exchange is offset in favour of
a relationship based more clearly on *mutual dependence* and hence on trust. Within the vertically-integrated firm, the existence of a set of bureaucratic rules determining the position of employees within the system of ranks, identifying the holders of managerial and supervisory power and regulating its exercise serve to create relations of *hierarchy*; the internal norms of the organisation are supported, in turn, by the external norms which grant legal force (up to a point) to employees’ expectations of job and income security and underpin their right to collective representation at the workplace. Rules of this kind, by institutionalising power, seek to render its exercise less personal and hence more predictable and legitimate (Ménard, 1994).

The elements of incompleteness and mutual dependence, power and trust, which are found in the contract of employment are also present in sub-contracting and franchising relationships, but not in the same form or, in each case, to the same degree. On the one hand, no inherent right of direction over production as a process, as opposed to control over non-human assets which may be the property of the customer or franchisor, is normally *implied* in these two cases. If the franchisor or customer wishes to exercise control over the form and process of production, the details must normally be specified in detail in the contract itself; but any such express contract, in addition to being costly to write and enforce, is inevitably going to be incomplete. The buyer will not have the implicit powers available to the employer to direct labour. Nor do subcontractors normally submit themselves to the direction of a single buyer in the manner that full-time employees do; customised subcontractors, despite entering into close and long-term relations with their buyers, nevertheless operate, or attempt to operate, some form of the ‘15% rule’ limiting their exposure to a particular customer (Lorenz, 1990). However, the existence of unilateral power in certain sub-contracting and franchising relationships is also recognised by some legal systems. While protective regulation of these relationships is slight in comparison to that of employment - there is very little regulation of franchising, for example, in most European countries - laws have been passed for franchisee protection in the USA, and in Japan the relationship between dependent subcontractors and their ‘parent’ buyers is the subject of statutory regulation.
More generally in inter-firm relations, the role of norms in achieving a balance between competition and cooperation in ‘industrial districts’ of vertically disintegrated firms seems crucial to the viability of successful productive systems. The cooperative aspects of the inter-firm relationships help to minimise the disadvantages of small firm size, while the competitive aspects, along with a high degree of specialisation, impart the dynamism and flexibility that are often lacking in large, integrated firms (Best, 1990; especially Chapters 7 and 8).

Marshall’s account of inter-firm organisation stressed the advantages of localisation of production, which include an increase in the degree and specialisation of skills and their diffusion throughout the community so as to create an abundant supply of appropriately qualified labour; the growth of ‘subsidiary’ trades and specialised services; and an increase in the use of highly specialised machinery made possible by the combined demand of many firms. The concentration of the member firms of an industry in close geographical proximity allows all to enjoy the benefits of large-scale industrial production and of technical and organisational innovation which are beyond the scope of any individual firm. Nevertheless, individualistic initiative and competition remained the central driving force of economic progress in Marshall’s analysis (Wilkinson and You, 1992). The importance of the localisation of production within industrial districts in Marshall is that it creates an environment more favourable to individual success. These positive external economies to individual firms stem, fundamentally, from their geographical proximity rather than any institutional structuring. The close proximity of firms within a particular industry provided opportunities for entrepreneurs to specialise and for the district as a whole to secure economies of scale (both static and dynamic) denied to isolated individual firms because of internal restrictions on growth. They can therefore afford to stay small and concentrate their initiative and inventiveness on what they do best. By doing so they create, in turn, an environment which improves the competitive position of the locality.

Marshall’s analysis has played a significant role in theorising about the success of local clusters of small firms in the north of Italy and elsewhere - so much so that they have been labelled Marshallian industrial districts.
However, contemporary analyses of industrial districts put much greater stress than did Marshall on the collectivist and institutional basis for successful coordination. Emphasis is placed on the influence of community - defined as family and other social relationships, rules of behaviour embedded in those relationships, and more formal institutions such as churches and political parties - in guaranteeing standards of behaviour which engender trust and cooperation and thereby strengthen inter-firm networks. Within industry, trade associations are seen as playing a central role in providing technical, financial, marketing, training and other services. Trade associations also represent employers in their dealings with local and central government and with organised labour. In turn, government establishes by social, company and other legislation a framework of standards which underpins the equitable and cooperative relationships between firms. Collective bargaining is important in establishing and maintaining effective standards of wages, employment conditions and training. Industrial and labour organisations play an active role in initiating and participating in the development of government industrial and labour market policy (Sengenberger, Loveman and Piore, 1990).

Important features of modern industrial districts are, therefore strong trade and labour associations; active industrial, labour market and social policies usually involving a role for representatives of the state, labour and capital; high standards for wages, conditions of employment and training; and close and direct relationships between firms. The outcome is seen as competitive success based on high rates of process and product innovation, the rapid diffusion of new techniques and good design, and high quality and wide variety in products (Brusco and Sabel, 1982). Moreover, this analysis is not confined to the industrial district model: effective normative governance of inter-firm relations involving (in various ways) industrial policy, legislation and trade associations also has an important role to play in the Japanese model of closely dependent buyer-customer relations and in the German model of ‘diversified quality production’ (Sako, 1992; Streeck, 1992). In different ways, the norms established by these means serve to minimise the risk to firms in trusting other firms and in entering into relationships with them which, to a greater and lesser degree, are based on an asymmetrical distribution of power. In promoting and enforcing such
general standards, the trade association or the state or possibly even a ‘parent’ firm or dominant buyer, as the case may be, effectively collectivises the risks which would otherwise be borne by individual firms and particularly heavily by the smaller ones, enabling these risks to be pooled among the group of firms as a whole.

5. **Institutional Origins of Trust**

The analysis presented above suggests, then, that the advantages of specialisation and an extensive division of labour rest on a prior normative framework which is capable of endowing trust, in part through controlling the abuse of contractual power. Economic theory has come to see trust as playing an essential role in underpinning contractual relations of various kinds (Gambetta, 1988); by contrast, the notion of power is treated with suspicion as being vague and malleable (Williamson, 1985: 238; 1993: 485). From a transaction cost perspective, trust reduces the information and monitoring which are inherent in complex principal-agent relationships and in court-based enforcement of agreements. The greater the degree to which relations in a productive system rest on trust, the lower will be both the costs of average material incentive (or threat) necessary to induce a given degree of cooperation and the average costs of monitoring and enforcing the incentive structure concerned. Power, on the other hand, carries a number of negative connotations: where there is a high level of asset specificity, the existence of sunk costs excludes, for a limited space and time, the disciplinary mechanism of the market. The result is a degree of private monopoly power which may be compatible with efficiency (Klein, Crawford and Alchian, 1978), although to what extent gains outweigh losses from the absence of competition is unclear in this approach. Alternatively, power may be seen not so much as a function of contractual monopoly as a corollary of the exercise of property rights (Hart and Moore, 1990), and on this basis can hardly be impugned since the individual ownership of property is a fundamental aspect of any system of exchange. Competitive markets are the best means of controlling the monopolistic abuse of power by property holders, through offering agents a wide choice of contracting partners.
This again neglects, however, the sense in which power offers a means of securing cooperation within contractual relations which are, to some degree, insulated from competitive forces by the parties’ joint need to realise the gains from a relation-specific investment. In this sense, power performs a role similar to trust in mitigating the effects of contractual uncertainty. In Luhmann’s early social systems theory, trust and power are seen as functional equivalents one or other of which may operate to minimise risk and complexity within social systems (Luhmann, 1979). Power enables the mutual expectations of the power holder and the other to be co-ordinated, with regard to the interests of the former. Adapting this idea to an economic context, we may say that power and trust are both means by which cooperation can be achieved at greater overall efficiency within a productive system, in terms of the reduced cost of putting incentives in place, monitoring performance and applying sanctions, and with greater potential benefits in terms of information flows and adaptation through innovation.

Power and trust nevertheless differ in important respects as means of responding to contractual uncertainty. While trust bears the positive connotation of a sharing of the rents to be gained from cooperation, with power it is more a question of a negative threat of a sanction being applied to gain compliance. In relationships based on power, moreover, the weaker party always has means available to counteract the threat posed by the power holder: transaction cost economics would recognise that asset specificity rarely if ever creates relationships of complete inequality. Employees may withdraw their cooperation through such tactics as a work to rule, or suppliers hold back on releasing information to a larger customer. Therefore, relations based on the exercise of power as opposed to the fostering of reciprocal trust inevitably carry with them higher monitoring costs designed to discover and penalise ‘defection’ by the weaker party. The benefits which flow from innovation and creativity based on the synergistic exchange of knowlege can be attained only with great difficulty in relations where trust is lacking.

Productive systems which are capable of generating trust therefore stand to gain an important competitive advantage over those which are not; but how does trust emerge and how is it maintained in the face of likely conflicts
of interest between autonomous contracting agents? It is possible to see trust as essentially calculative in nature. Calculative trust arises from the strategic interaction of self-interested economic agents and, as explained above, is associated with the contention of game theory that long-lasting, self-enforcing agreements are possible within the framework of repeated trades. Parties remain in such relationships as long as they regard it as being in their self-interest to do so, the possibility of defection being restrained by the sanction or retaliation which would then be deployed against them (Dasgupta, 1988). A rational-choice framework is thereby provided to explain institutional or cultural factors which are associated with trust, such as corporate culture (Kreps, 1990; see more generally Coleman, 1990).

But there would seem to be an inherent contradiction in a notion of trust as stemming from calculative behaviour. At the very least, it would seem that such a notion of ‘trust’ adds nothing of significance to the analysis. It would be strictly unnecessary to speak of trust if it were merely the expression or outcome, however elongated, of private self-interest coupled with the efficient choice of contractual incentive structures. If the examples game theory gives of cultural or institutional forms which are said to be indicative of trust can well be explained in choice-theoretic terms, then the notion of trust as the foundation of cooperation is either redundant or just confusing.

An alternative approach is to stress the inter-personal nature of trust. ‘Personal trust’ is based on close, long-term relations, a sharing of goals and expectations and the suppression of short-term self-seeking. This can result in a degree of altruistic behaviour: there may be a willingness to forgive and forget rather than to retaliate in the case of a breach of promise, or a refraining from taking advantage of a short-term opportunity, in each case reinforcing the parties’ mutual expectations that their relationship can and will be maintained in the future. In this approach, the parties’ behaviour is not reducible to self-interest, even of a long-term kind: it is suggested instead that social or personal trust arises from shared cultural values and history, the development of friendships, and gift-exchange (Lyons and Mehta, 1994). An example of this is the ‘goodwill trust’
identified by Sako (1992) in the context of Japanese-style buyer-supplier relations, and which she contrasts to the more limited ‘contractual’ trust which is said to characterise relations between British subcontractors and their customers. Goodwill trust of this kind is seen as being inimical to the use of detailed and formal contractual agreements involving risk-allocation mechanisms and recourse to law (Lyons and Mehta, 1994).

However, while the existence of personal trust may well facilitate certain forms of cooperation in long-term business relations, such as those in family-run firms, these are exceptional and, even then, may be characterised as much by differential power and distrust of a personal kind as by its opposite. The clear opposition between trust and law, which is inherent in the view of trust as personal or cultural, is also questionable, at least in a European business context (Lane and Bachmann, 1995). The fact that the parties to contracts retain clear and separate interests of their own means that in many cases they will have resort to formal, legally-binding agreements as a way of handling residual risks, and instances in which contracting parties actively seek to avoid placing their agreements in some kind of legal form are rare. Although both parties know that going to court to resolve a dispute will effectively end their relationship, a formal contract specifying rights and obligations in the event of certain contingencies can be used as a means of averting this possibility. Most short-term or one-off trades involve the use of order-forms containing legally significant terms and conditions, and long-term supply arrangements involving single sourcing are normally put into contractual form following bilateral negotiation; entering into such a contract may well be a sign of an already-existing relationship in which founded at least partially on trust. What is important here is that the use of contracts does not necessarily imply an inferior form of trust by comparison to the more diffuse and open-ended nature of Japanese supplier-buyer relations, as Sako (1992) implies: the elements of goodwill and personal trust frequently co-exist alongside the more formal elements supplied by the use of contractual documents and the existence, in the background, of the legal system of contract enforcement.

A clearer answer to these questions may be obtained by focusing on the
institutional origins of trust (Deakin, Lane and Wilkinson, 1994; Lane and Bachmann, 1995). Luhmann (1979: 40), for example, looks for 'the foundations of trust in the social order'. Law is one important mechanism for dealing with the essential riskiness of trust: 'legal arrangements which lend special assurance to particular expectations and make them sanctionable... lessen the risk of conferring trust' (Luhmann, 1979: 34). Similarly, Zucker suggests that even trust of a personal or cultural kind - which she defines as 'process-based', in the sense of arising out of long-standing relationships, or 'characteristic-based', that is to say resting on common family, ethnic or religious characteristics - needs to be supported, in modern economies, by a form of trust which is rooted in stable institutions. This 'institutionally-based' trust operates when trust is tied to formal, societal structures which have an existence separate from the material preferences, motivation and actions of individuals. Institutional forms therefore are not reducible to rational choice considerations, but must be understood as playing an independent role in structuring individual agency.

6. Institutions, Organisational Form and Competitive Advantage

Cooperation in relational contracts, whether intra- or inter-firm, therefore rests on the presence, beyond the contract, of organisational forms and institutional regulation of a certain kind. These institutional forms are more or less rigid and operate to a large degree independently of the motivations and behavioural traits of the economic actors themselves. Institutions and organisations are distinguishable from each other in terms of the degree of rigidity or permanence attached to them as sources of contractual governance. Institutional sources of norms include the legal system, the mechanisms of economic management directly available to the state (such as taxation, control of public spending, and industrial and macro-economic policy), and also non-state bodies of various kinds including trade associations: these possess a high degree of rigidity and stability through time. Organisations, including firms, network relations and also markets themselves, are less stable and operate within the general framework set by institutional norms, while themselves also operating as structures for the governance of exchange. An understanding of the
processes determining the evolution of these forms will be at least as important to an appreciation of the success of productive systems as an understanding of the logic of market competition.

Changing laws and regulatory forms is possible, but costly; changing organisational forms is also costly, although less so. In this world, the firm-market boundary will not be set by the efficient contractual choices of economic agents, as suggested by principal-agent and property rights theory, nor by the ‘moving equilibrium’ Coase identifies in terms of the comparison between the marginal costs of organisation on the one hand and decentralised exchange on the other. Instead, the manner in which production is organised will depend both upon the strategic choices of a firms’ management and upon the state of the contractual environment in which it operates, that is to say the state not just of available technology and information but also the organisational and institutional structures for governing exchange.

It follows that the degree of uncertainty or of asset specificity in a given transaction - even assuming that these can be accurately measured, which may be doubtful - will not be a definitive guide to the form in which a productive system is constituted. In particular, it is not possible to say that contractual relations are arranged along a spectrum according to degrees of asset specificity, with the firm at one end, the market at the other and ‘hybrid’ relations in between. Although asset specificity may be an essential concept in explaining the breakdown of decentralised market relations, its use to explain differences, for example, between the vertically integrated firm and the use of customised subcontracting is much more problematic. This is because the option of using the latter may be increased or decreased by a factor which is independent of the degree of specificity of a particular asset, namely the existence (or otherwise) an institutional framework for inter-firm relations of the kinds discussed above.

For some, it will no doubt remain axiomatic that economic development proceeds on the basis of the ‘machinery of spontaneous coordination’ of market relations (Robertson, 1928: 87), at the heart of which lies a minimal system of rules embodying Hume’s three fundamental laws of nature:
stability of possession, the right of disposition of property by consent and the performance of contractual promises (Hayek, 1988: 34). From this point of view, collective norms designed to regulate contractual power and impose quality standards on market entrants are instances of ‘restraint of trade’ and inter-firm cooperation a species of ‘collusion’. Institutional forms (including law) should themselves be subjected to competition in order to ensure the survival of the fittest through an evolutionary process, akin to natural selection: in the long run those which are most effective (if not necessarily optimal) will survive and prevail (Posner, 1986).

But if institutional forms have a certain degree of rigidity, their evolution may be far from efficient; sunk costs mean that institutions, just like firms, may become too highly specialised to adapt to changing external circumstances (North, 1994). The rigidity through which institutions impart necessary predictability and stability to economic systems may turn out later to be a cause of institutional failure when technologies change or there are shifts in the composition of demand. This is not to say that conscious policy making can necessarily reverse a process of institutional failure and economic decline, even after it has been clearly identified. However, conscious policy making has unquestionably contributed, in some cases, to the development of institutional structures which have proved more successful than others. State intervention does not have to take the form of ‘planning’ or ‘command’ in the sense of direct state ownership or direction of resources. However, supporting trade associations and instituting procurement practices designed to foster long-term relations and the promotion of quality are devices which may be used to provide underpinning for inter-firm relations, just as employment protection and collective bargaining or co-determination laws may assist the growth of cooperative industrial relations within firms. These are just a few examples of practices which tend to be justified in practice by reference to considerations of ‘equity’ or ‘solidarity’, but which may equally well have important efficiency-enhancing properties.
Notes

1. For example, in the case of knitwear the vertical linkages would include the production of fibre, the spinning of fibres into yarn, the knitting of yarn into fabric, the manufacture of garments and their marketing. The horizontal links would include the production of machines and other means of production, the training of labour and the provision of product design and other services specific to each stage.

2. For a more detailed discussion see Tarling and Wilkinson (1987).

3. For these purposes we are not attempting to distinguish between risk (the degree of which can be calculated) and uncertainty (which cannot) in the manner of Knight (1921).

4. Arguably this view is implicit in Coase’s own 1937 paper, which sees the limits of the firm as being set by the degree to which ‘it pay[s] to bring an extra exchange transaction within the organising authority’ (ibid.); the extent of vertical integration depends on the comparative marginal costs of organising exchange within one framework or the other. An equilibrium framework is thereby imported back into the analysis.

5. At least, this is said to be true of the common law of contract. Whether it holds for the civilian tradition of commercial law is more doubtful (Lane and Bachmann, 1995).

6. For example, two people lifting an object can together lift more than the sum of what each can lift separately. For other examples of simple cooperation see Wakefield (1835) and Marx (1974: Chapter XIII).

7. For details of the benefits of simple cooperation see Marx (1974, Chapter XIII).

9. Marx also recognised the importance of the coordinating role of capitalists: ‘A single violin player is his own conductor; an orchestra requires a separate one. The work of directing, superintending and adjusting becomes one of the functions of capital, from the moment that the labour comes under the control of capital [and] becomes cooperative’ (Marx 1974: 313).

10. A term Marshall preferred to ‘competition’ because of the need for a term ‘that does not imply any moral quality, whether good or evil, but which indicates the undisputed fact that modern business and industry are characterised by more self-reliant habits, more forethought, more deliberate and free choice’ (1947: 9-10).

11. This part draws on joint work with our colleagues Reinhard Bachmann and Christel Lane. See Bachmann (1994), Lane and Bachmann (1995).

12. For the development of this distinction in an extended transaction cost framework, see North, 1990 and 1994; ‘if institutions are the rules of the game, organisations, and their entrepreneurs, are the players’ (1994: 316). New institutionalist theory in sociology has developed in greater detail a framework for analysing the state and its role in the economy: Powell and Di Maggio, 1991; Whitley, 1992; Lane, 1994. These approaches have developed separately from one another but have many features in common, most notably their recognition that institutions have an ontological status in their own right.
References


Kreps, D. 1990. ‘Corporate culture and economic theory’. In J. Alt and K. Shepsle (eds.) *Perspectives on Positive Political Economy*. Cambridge:


Lane, C. and Bachmann, R. 1995. ‘Risk, trust and power. The social constitution of supplier relations in Britain and Germany’. Working Paper 5, ESRC Centre for Business Research, University of Cambridge.


Lyons, B. and Mehta, J. ‘Contracts, opportunism and trust’. Working Paper, School of Economic and Social Studies, University of East Anglia.


Zeitlin, J. 1994. ‘Why are there no industrial districts in Britain?’ Mimeo, University of Madison-Wisconsin.