Journal of Institutional Economics (2015), 11: 2, 245–264
 doi:10.1017/S1744137414000265
First published online 26 June 2014

Ronald Coase's theory of the firm and the scope of economics

BRIAN J. LOASBY*
Division of Economics, University of Stirling, Stirling FK9 4LA, UK

Abstract: Ronald Coase's work and its reception illustrate the significance – and the difficulty – of identifying problems and proposing solutions, which provides the theme of this paper. His theoretical innovation was not derived from economics, and seemed irrelevant to contemporary issues of economic theory and policy; only his much later perception of an apparently unrelated problem – the incoherent treatment of social cost as market failure – showed how the concept of transaction costs could illuminate two major areas of economics. The inadequate treatment by economists of the transaction costs of markets is linked to the neglect of processes, and especially the processes of organising the growth and use of knowledge – key concerns of Smith and Marshall. The curious relationship between Coase's explanation of firms and Austin Robinson's analysis of competitive industry leads to a reflection on the scarce resource of human cognition and the role (and fallibility) of institutions.

1. Prologue

We try to make sense – and sense has to be made – by inventing patterns and imposing them on selected phenomena and situations. Every system that we construct has a limited range of convenience, but the limits can never be known for certain, and may be unsuspected until we exceed them. Interpretative failure may then provoke the search for new patterns which seem to work where the old patterns do not; but we cannot reliably predict when this will happen, still less the content of our new knowledge, or the consequences of applying it. What we often find is that the acceptance of new patterns which appear to resolve significant difficulties is impeded by conflicts with well-established systems (especially those of wider scope such as universal theories of economic systems or established principles of running a business) which continue to provide comfort in other contexts that seem more important to us. This conception of the incentives and obstacles to the development of knowledge offers a valuable insight into the problems of change in both scientific and economic systems.

*Email: b.j.loasby@stir.ac.uk

2. Firms

Ronald Coase's work supplies an important example of this obstacle to inventing, accepting and applying new patterns. As he himself observed, 'the firm' was an unexplained element in a theory which demonstrated how economic systems were strictly regulated by markets that were governed by the basic data of preferences, resources and production possibilities. For Lionel Robbins, this was what mattered, and, as Coase (1991b: 53) noted, he deplored the waste of Marshall's abilities on the study of business practices. However Coase escaped Robbins' theoretical influence because he lacked the school qualifications required for the LSE economics programme, and took a commerce degree, where in two years he 'studied a great variety of subjects, devoting very little time to each and inevitably doing no systematic reasoning' (Coase, 1991a: 36–7). He therefore had no exposure to formal price theory: many years later, he observed that 'in 1932, when I formulated my ideas in "The Nature of the Firm", my analytical system, such as it was, came from [Arnold] Plant' (Coase, 1991b: 49). Plant, who had been appointed as a professor of commerce in 1930, was no less insistent than Robbins that competition delivered co-ordination, and Coase 'did not dispute what Plant was saying. Indeed it made a lot of sense. But it seemed somehow incomplete' (Coase, 1991a: 38).

However, a sense of incompleteness is not sufficient to produce a new idea. Indeed all theories are incomplete, relying on omissions and assumptions which are seem either reasonable or (perhaps too often) convenient, or even indispensable. What set Coase on a path to a solution was Plant's reference to 'the different ways in which firms were organized' (Coase, 1991a: 38); these differences provided a theme for a practical study of vertical and lateral integration in the USA, which was financed by a travelling scholarship for the academic year 1931–2. Trying to understand what is happening without a guiding theoretical concept is not usually recommended, but for Coase it worked.

What was ironically called 'the theory of the firm' could give no theoretical reason for the existence of firms, because it relied entirely on market transactions to explain the prices and quantities of all goods and services. This theory simply required consumers and producers, all conceived as individual agents: in the goods market consumers provided the demand curve and producers the supply curve, and in the labour market the roles were reversed. Demand curves were conceived to be directly derived from individual preferences, which were subjective but well-ordered, and supply curves from costs, which were determined by technology and resources; and preferences, technology and resources were all presumed to be objective data. The intersection of these curves, properly defined, was then sufficient to determine outcomes; there was no need to explore market processes. As Coase (1988: 28–30) observed, these assumptions legitimized reliance on 'blackboard economics' for both analysis

and policy prescriptions. Only someone who, like Coase, had studied business practices but had never taken a course in price theory, could possibly conceive the need to provide an existence proof for firms, let alone to investigate what firms do.

For those who were familiar with the apparatus of price theory Coase's existence proof therefore did not seem either to suggest any modification to that theory or to define a new field for research in economics. I suggest that this was why Lionel Robbins never discussed Coase's article with him, despite their cordial relations. Robbins had already celebrated the absorption of production into the theory of equilibrium, and business practices had no causal significance for understanding economic outcomes, because it was clearly understood that no individual or group could escape the logic of the market.

For Robbins, this logic required every firm to produce the level of output at which price was equal to marginal cost, and so a market economy was inherently efficient. In the 1930s this conclusion was seriously challenged by the proposition that the market system necessarily produced serious distortions of prices (including wages) and quantities. The assumption of a widespread tendency for cost to be a declining function of output – which was at least in part the result of representing processes by static models – compelled firms to operate at equilibria in which marginal cost was equated with marginal revenue, which was well below price – at least in the diagrams, which was sufficient to establish the result. The advocates of this new microtheory were no less convinced than believers in efficient markets that the survival of every firm required conformity to market imperatives, and concluded that since these misallocations were dictated by the market structure they could be removed only by changing that structure (which might well be impossible) or by replacing it with some form of planning, which simply required the application of logic to the data. This analysis appeared to demonstrate the power of economic reasoning to make a major contribution to human welfare. However nowhere in this argument was there any recognition of the need to compare the cost of market transactions with those of internal management, which was Coase's conceptual innovation, or with the cost of planning: presumably none of these costs would be incurred in the equilibria being analysed. So Coase's argument, though ingenious, seemed irrelevant to the important analytical and policy issues of the time.

This irrelevance is clearly illustrated by James Meade's (1936) proposed remedy for the substantial welfare losses which were necessarily imposed by imperfect competition. The set of products in each industry should be centrally specified, and in each firm within that industry salaried managers should be instructed to choose the level of output at which marginal cost was equal to price. As a reasonable man Meade was duly concerned about the loss of the profit incentive in this novel regime, but he observed that, whereas in order to maximize profits in imperfect competition entrepreneurs needed to know both marginal cost and marginal revenue, the managers in his proposed system

needed only the former; moreover the calculation of marginal cost required only internal data which was readily available. He apparently knew nothing of accounting or, more fundamentally, of the complexities of production systems. Thus the reduced incentive would be offset by substantial simplification of the task, which itself was quite straightforward. All transactions, in firms as well as markets, were implicitly assumed to be costless – in accordance with standard practice. Thus what actually goes on inside firms was not an interesting question for either Meade or Robbins – although for different reasons. Coase, of course, was repeatedly scathing about the confident ignorance of economists when pronouncing on public policy.

It is clear from the record that the existence of firms as economic agents seemed irrelevant to the core of microeconomic analysis during what Shackle (1967) later called 'the years of high theory'. In Keynes' macroanalysis their role was simply as issuers of bonds and shares the future value of which was necessarily unknowable; but no-one was interested in such unpredictability in microtheory, for which economists already had a universal principle of theory-building with which they were comfortable, even though they might disagree sharply about particular applications of this principle to economic policy. Most of them were willing to enshrine as an ideal system a concept of perfect competition which, as George Richardson later observed, 'might reasonably be regarded as a denial of Smith's central principle [of economic progress through the interaction of competition and the division of labour] erected into a system of political economy' (Richardson, 1975: 353), and then to enter into a spirited debate on the implications of imperfectly competitive equilibrium. Marshall's warning of the limitations of equilibrium analysis and his careful definition of increasing return as a continuing process of organizational change and improving knowledge, in which equilibria are necessarily, and desirably, local and temporary, were dismissed as the result of timidity and confusion respectively.

Sraffa's (1926) reduction of increasing return to a characteristic of a production function, because this was a necessary requirement of equilibrium analysis, and in particular of his own theoretical programme, seemed impervious to Allyn Young's (1928) eloquent elaboration, two years later, of the central role of increasing return as a creative process driving economic development. Equilibrium was apparently indispensable for economic reasoning, and this reasoning had striking implications for both theory and policy (as illustrated by Meade's book). That Coase's theory was 'much cited and little used' (Coase, [1972] 1988; 62) – and is still often misunderstood – is therefore both scandalous and easy to explain. What had Coase to contribute to the central debate of the time?

May I add a personal recollection? As an undergraduate at Cambridge I was repeatedly exposed to spirited (and not always well-tempered) arguments about Keynesianism and microtheory. Three years as a member of the Economics Faculty in this environment had been quite enough for John Hicks (Creedy, 2013:

225). There was an assumption, shared by most protagonists in Cambridge, that those with Cambridge connections were particularly worthy of notice, but outsiders were not necessarily excluded (though I do remember that Hayek emphatically was – an exclusion which I did not then challenge). Philip Andrews did receive some attention, but that was because he was regarded as a misguided, but potentially dangerous, opponent of imperfect competition theory. His claims that his arguments were based on his investigations of firms (Andrews, 1949) elicited the response that he had either misunderstood what businessmen had told him or been deliberately misled by them; the theory was logically unassailable. I can also recall hearing the suggestion that the businessmen who talked to Andrews might actually behave in the ways that they described, and that this clearly demonstrated their incompetence, which of course aggravated the necessary inefficiency of imperfect competition. I was already inclined to believe that Andrews understood what happened in firms rather better than his Cambridge critics, and I may say that reading Meade contributed to my increasing undergraduate scepticism about the relevance of the new price theory.

This experience may suggest that I would have been receptive to Coase's argument had I known about it at the time – which I didn't because no-one mentioned it and we didn't read back numbers of journals. It was not until the republication of his paper in a set of readings (Stigler and Boulding, 1953), that I discovered it. However I should admit that it was not until I began thinking about the present paper that I recognized a connection between the interests and approach of Coase and Andrews – of which neither of them apparently ever became aware. (From this I draw two morals: first, it's never too late to learn, and second, making appropriate connections is not straightforward, and has nothing to do with logic, although it is a precondition of useful logic.) I was already more interested in business behaviour, and in tracing the development of an industrial district, than economic theory, and Coase's first paper was concerned with the possibility of reducing transaction costs by creating regions of managerial discretion which (adapting Shackle's (1972: 160) characterization of money) is a 'means by which choice can be deferred until a later and better-informed time'. However Coase was not then investigating how such regions might be constructed or chosen and how this discretion should be exercised – what we might, anachronistically, call the 'Carnegie questions' associated with Simon, Cyert and March – which became my focus. So it took me a long time to perceive its relevance for my own concerns.

Coase himself was not quick to explore the field which he had opened up. Indeed his next substantial theoretical contribution, which came 23 years later and eventually assured his Nobel Prize, did not examine what happens inside firms but applied the logic of transaction costs to the problem of unpriced externalities in economic activities, and concluded that in a world without transaction costs there is no reason why any externalities should be unpriced, and therefore that the established corpus of welfare economics, in which transaction

costs were ignored, was not a sound basis for economic policy. This argument received immediate attention because it was interpreted as a contribution to a prominent debate about allocative efficiency, although what came to be called 'the Coase Theorem' misrepresents Coase's unwelcome conclusion that there is no adequate substitute for case-by-case investigation.

It was another ten years before his address to the National Bureau of Economic Research in 1970 marked a shift in his focus from the choice between market and internal transactions to the distribution of activities between firms, although this had always been on his agenda. This shift was accomplished by a simple extension of his framework to incorporate differences between firms in the costs which they would incur by internalising particular transactions, leading to the crucial observation that the economies which may be gained by bringing together classes of transactions which are similar in important respects provide a means of explaining the scope of each firm's activities. (Coase, [1972] 1988: 63). However, his focus remained on the allocative efficiency of combining related activities rather than the development of increased skill in performing them, which Marshall and Penrose had emphasized.

I suspect that Coase was right to regret his own earlier emphasis on the contractual arrangements for the supply of factor services, at the expense of considering the management of these services (about which he had learned a good deal in America), because this encouraged economists to 'neglect the main activity of a firm, running a business' (Coase, 1991b: 65). That he failed to cite either Andrews or Penrose as important exceptions to this neglect illustrates the difficulty of making connections. Running a business, of course, involves far more than decisions about how to use the discretionary powers which have been specified in the contract of employment. That such specifications may be very imprecise – or even non-existent – is not the main problem; indeed imprecision may be a partial solution to the main problem, which is how to decide what to do next and how to do it.

This question was a fundamental concern of Shackle, in both macroeconomics and microeconomics (as well as the history of the discipline). Although well aware that he had no specific aptitude for business, Shackle was fascinated by the problems and processes of management, which (like Knight) he perceived as a continuing struggle, not only to cope with uncertainty, but also – and especially – to exploit uncertainty by developing new products and new ways of working. In his theory of the firm, commissioned by Charles Carter (who was also interested in business practice), Shackle (1970) offers no long-run equilibrium for individual firms; nor, as is rarely recognized, did Marshall. (For an unexploited connection between Shackle and Coase, resulting from their separate interactions with Henry Boettinger, A T and T's Director of Corporate Planning, see Loasby, 2011: 773). We shall return to the problems of running a business, and the implications for the choice between internal and external relations; but in order to consider them properly it is first necessary to follow Coase's logic a little further.

3. Markets

Coase's proposition that firms, which operate substantially through general authority rather than specific contracts, often provide lower-cost means of resource allocation than markets may be justifiably regarded as an ingenious extension of marginal analysis to the choice between external trade and internal direction. However it relies crucially on the assumption that the use of markets normally incurs costs; and in relation to standard price theory what matters is that these are not production costs, which are included in the determination of equilibrium, but costs which are incurred by agents in discovering equilibrium. Coase ([1960] 1988: 114) summarises the sources of these costs: 'it is necessary to discover who it is that one wishes to deal with, to inform people that one wishes to deal and on what terms, to conduct negotiations leading up to a bargain, to draw up the contract, to undertake the inspection needed to make sure that the terms of the contract are being observed, and so on'. Now if one is seeking the highest standard of theory this distinction between production and transaction costs is crucial. The Arrow-Debreu model achieves completeness by requiring every good to be defined, in addition to its technical specification, by location, date, and the state of the world at that date. In order to close the model it is necessary for the economy to terminate – forever – at the latest date which is included. This specification, together with a complete preference function for all consumers, allows a proof of general equilibrium. That this equilibrium may, in principle, be attained by market transactions, or by a planning commission using the same data about preferences, resources and technical possibilities, was regarded by some prominent economists as a valuable guide to policy.

However there are two problems. First, if the equilibrium is to be achieved by trial and error, in which every agent attempts to conclude the best possible set of transactions, then it is essential that no actual production or exchange occurs until the full equilibrium is established, because, as Walras had discovered, any production or exchange at what eventually prove to be non-equilibrium prices has the dual effects of redistributing income between agents, thus changing their opportunity sets, and applying some productive resources in suboptimal combinations and to uses which turn out to be inappropriate. These effects change the outcome from that which is implicit in the data; so neither predictability nor optimality can be expected. So, as good general equilibrium theorists have pointed out, the economy must not open until the market has closed. That may suffice as a formal solution to the formal problem, but since everyone has already agreed what to do in every possible future situation before the economy opens, an operating economy incurs no transaction costs and therefore, by Coase's reasoning, no firms – as indeed is true of the Arrow-Debreu model, which is populated by individual agents. We may, however, think it odd that the actual operation of a supposedly ideal market economy makes no use of markets.

The second problem is that the market activity which necessarily precedes any production or exchange must use none of the resources which are to be allocated in the equilibrium: if there are costs of finding equilibrium these must be qualitatively different from the costs of a functioning economic system. Now we may not be very concerned about the content or uses of high theory; but we should be concerned about the costs of running any actual system, in which transaction costs are very substantial. We may also, like Coase, prefer theories which give credible explanations of real phenomena. Coase has given us credible explanations: firms may develop organizational structures which are particularly effective in managing specific classes of problems, though he does not investigate how this happens or the implications for the structure of industry. Indeed his 'Proposal for Research' (Coase, [1972] 1988) emphasizes the current lack of knowledge and argues for an extensive investigation of the range of activities undertaken by American firms, together with their contractual arrangements, to provide a basis for theorizing. He was clearly unaware of the basis which had by then been provided by Penrose (1959, 1995), as she seems to have been unaware of his work when producing her own .

Coase's explanation for the emergence of firms raises new questions. This is not unusual; indeed the tendency of solutions to raise new problems is an important driver of the growth of knowledge, in both economic and scientific systems. Now if we accept Coase's basic proposition that there are costs of using markets – which is obviously true, and that relying on internal relationships may often achieve similar results at lower costs (although still leaving us with a multitude of markets) then we should perhaps begin by asking who bears the costs of markets and why. In fact should we not start by asking how markets come into existence?

Simple answers have been given to this question by people who were dealing with other questions, but – significantly – always within a moving economy; markets are needed only if there are contracts still to be made. We may distinguish three such questions, with corresponding answers. The first, already noted, is the basic problem of attaining an equilibrium which is inherent in the data. Here the basic notion is that if an economy is not already in equilibrium, then movement towards equilibrium generates gains which may be (in part) captured by alert individuals; this prospect may therefore motivate these individuals to incur some costs in order to attain them; these may include the costs of creating and operating markets which are dedicated to this purpose. It is not assumed that an equilibrium which is attained by this process is identical to that which could 'in principle' be deduced from the data – where 'in principle' means 'impossible'.

The simplest version of this explanation, although not quite presented in this way, is Kirzner's (1973) theory of entrepreneurship. Kirzner's basic assumption is that circumstances change, because of external events, chance discoveries, or individual initiative. These changes generate price differentials between locations, or differences between the prices of productive inputs and their enhanced value

in new uses. Such differences create disequilibria; however, instead of simply assuming that markets will eliminate them (perhaps with an explicit assumption of perfect information which, as Coase (1991b: 69) noted of a similar ploy, puts the rabbit into the hat from which it is subsequently produced), Kirzner recognizes that opportunities have to be noticed and that different people tend to notice different things. This differentiation matters, because responding to an opportunity incurs a cost, and so there must be a prospect that the opportunity will last long enough to justify incurring this cost. As George Richardson, enquiring how an economy could reach equilibrium, observed, '[a] general profit opportunity . . . is . . . a profit opportunity for no one in particular' (Richardson, 1960: 57). Kirzner envisages the creation of particular markets to allow their creators to benefit from their differentiated knowledge; securing these benefits distributes this knowledge and so creates a new equilibrium in which each particular market disappears.

However the differentiation of knowledge, although necessary for such processes, is not sufficient. Specialisation between the stages of long production chains, as envisaged by Menger (1976), may be necessary because goods which are complementary may require very different capabilities – the context-relevant practical skills which are not recognized in the standard concept of a production function; and the consequent differences between mindsets are liable to impede both the perception of opportunities for new combinations and the effective management of responses. Inventing a 'representative agent' eliminates the theoretical problem – for blackboard economies; but the incentive to bear the costs of creating a particular market requires at the least differential and localized alertness, and sometimes much more. Instead of being perceived, opportunities may need to be created, essentially by conceiving new patterns or transferring established patterns to new contexts.

In the conventional language of economics we therefore move from the perception of disequilibrium to the deliberate disturbance of an existing equilibrium, and from Kirzner to Casson. Casson's (1982, 2003) entrepreneur conceives a new product, and is prepared to bear the costs of developing a new market for it. If the new product is to be produced by a new business (which is Casson's basic case), then the entrepreneur is faced with precisely Coase's problem: how to choose between external sourcing and internal management for every part of the production and marketing process. Casson's entrepreneur uses both; and a crucial factor in making this selection may be doubts about the capacity of potential suppliers or distributors to understand precisely what is required – or (as with Henry Ford's failure to induce his suppliers to introduce assembly-line production) their willingness to do it.

The third answer takes us from Casson's localized innovation to Schumpeter's large-scale 'creative destruction'. The transformational innovator, although benefitting from the low transaction costs of a well-established pattern of relationships – this is a key element in Schumpeter's model – must be

willing to make substantial investments in what we may call the transaction costs of transition to new transaction cost regimes. Such willingness requires rare personal characteristics. However the success of such entrepreneurship necessarily entails the wholesale invalidation of many market relationships and internal arrangements on which existing agents have come to rely, and for which substitutes are difficult to create. Schumpeter's paradox is that the initiation of large-scale entrepreneurship requires pervasive equilibrium and relatively low transaction costs in the construction of new combinations, but necessarily destroys that equilibrium and in so doing raises transaction costs in both firms and markets in large sections of the economy, so reducing the level of economic activity. For Schumpeter recession is a necessary element in economic progress.

I shall not pursue that theme here. However, what I shall take from it are two principles. The first is that economists should pay much more attention to processes – both the economic processes of co-ordination and innovation which the acceptance of particular kinds of transaction costs make possible and the processes of changing the costs of those transactions which seem important (as judged by both insiders and outsiders): this is typically an important element in entrepreneurship. Changes of costs are often achieved by modifying the forms of these transactions, as we shall see. Coase's arguments about the firm and the law often draw on the importance of experience and adaptation, and firms and markets (and many intermediate forms of relationships) have particular effects in guiding processes. The particular processes which I now propose to consider briefly are those which help to shape the growth, transmission and application of knowledge. The second principle is that we should think of systems as quasi-decomposable; indeed it is hardly possible to understand processes in any other way – as the history of general equilibrium theory demonstrates.

4. Organization and knowledge

The basic ideas in this section come from Adam Smith's contemplation of human characteristics. First is the desire for mental comfort, which provokes the creation of patterns which seem to accommodate troubling phenomena within some kind of order (Smith, [1795] 1980). Next is the interest that we take in other people, and in particular our willingness to copy apparently successful behaviour which we observe in others and to engage in some co-operative activities. These characteristics, which substantially reduce transaction costs, we may see as a psychological precondition for firms, and for the willingness to make selective exchanges of goods and services, which, as Adam Smith ([1759] 1976a) perceived, is the psychological precondition for markets. A narrow definition of self-interest cannot support an adequate theory of a market economy. The combination of these two propensities leads people in different situations to give selective and more detailed attention to a wider range of phenomena, and by generating novel problems this differentiation provides

individual incentives to develop more varied and more detailed schemes of order, each of which seems appropriate to a particular context – a propensity which may be hard to incorporate in any theory which is based on a 'representative agent'. It should be noted that Smith's theory of the nature and growth of human knowledge respects Hume's twin warnings that logic can never produce a novel idea and that no amount of evidence can establish the truth of any universal proposition.

These ideas are brought together in the *Wealth of Nations* to produce a powerful theory of continuing growth through decomposition and new combinations: a finer division of labour generates new ideas for new contexts – a theme already incorporated in Smith's ([1795] 1980) theory of the growth of knowledge, and the application of these ideas increases both the variety and volume of goods and services, thus permitting an even finer division of labour. This process can be self-sustaining (Smith, [1776] 1976b), and justifies both the acceptance of transaction costs and the search for ways of reducing these costs. Knowledge and organization are thus doubly linked: knowledge itself is produced by making specific connections within the brain, as explained much later in Hayek's (1952) theory of the mind, and the patterns of social and economic relationships influence the kinds of knowledge that emerge.

I have become increasingly convinced that this interaction between organization and the growth of knowledge is a key – even perhaps *the* key – to understanding Marshall's *Principles* (Marshall, 1920), and not least to understanding why he insisted on partial equilibrium, the significance of which Samuelson totally failed to grasp because of his exclusive focus on allocative efficiency – even when modelling growth. Indeed, this seems fairly obvious if we remember, first, Marshall's own explanation that his fundamental reason for choosing economics as his field of study was his desire to understand how to improve the conditions of the people, for which, in the long run, improved allocative efficiency and reallocation within a given set of goods and technological possibilities were clearly insufficient – as they still are, and second, that the *Principles* and *Industry and Trade* were intended to be elements of a single scheme to improve this understanding. This concern with continuity and change disappeared after 1926, when Sraffa transformed increasing returns from a crucial driver of continuous improvement into a threat to efficient allocation within a closed system.

Marshall's orientation is reflected in his initial discussion of the agents of production. 'Capital consists in great part of knowledge and organization: and of this some part is private property and other part is not. Knowledge is our most powerful engine of production . . . Organization aids knowledge; it has many forms, e.g. that of a single business, that of various businesses in the same trade, that of various trades relatively to each other, and that of the State providing security for all and help for many' (Marshall, 1920: 138–9). This conception is not well represented by the standard model of aggregate growth

https://doi.org/10.1017/S1744137414000265 Published online by Cambridge University Press

theory; by contrast Marshall's suggestion that 'it seems best sometimes to reckon Organization apart as a distinct agent of production' (Marshall, 1920: 139), especially when this is linked with its 'many forms', now seems to invite Coase's enquiry into the basis for choosing between firms and markets as organizational forms and his invocation of their relative costs to provide this basis.

Marshall's complex system played no part in stimulating Coase's work. However we might ask why Marshall himself did not do something similar. We may at once notice that Marshall was well aware that using, and indeed organizing, markets entails costs, and also that he gives us a clear indication of who bears them, and why. In his extensive discussion of marketing in *Industry and Trade* (which, not surprisingly, Coase had not read when formulating his ideas) Marshall observes that in product markets the costs are normally borne by sellers and in labour markets by buyers, because these are usually the volume traders and therefore have both the greater incentive to facilitate efficient exchange, especially when introducing new products, which is essential to Marshall's concept of competition – remember his citation (Marshall, 1920, 280) of Roscher on the creation of new wants as a characteristic task of the modern manufacturer – and the greater scope for achieving scale economies in doing so (Marshall, 1919: 271–4). He also observes that firms generally distinguish between their 'general' and 'particular' markets (Marshall, 1919: 182); these have different transaction costs. What, I suggest, prevented him from anticipating Coase is that he seems never to have accepted – or perhaps even considered – Coase's clear, and crucial, distinction between exchange and direction as the principles on which markets and firms operate. The relationships between 'businesses in the same trade' and between 'various trades in relation to each other' are clearly not intended to be thought of as purely market relationships; if they were, then Marshall would not have referred to them as 'forms of organization'.

Richardson (1972) provides examples (many drawn from his work for the Monopolies Commission) of the complexities of some of these relationships, and also provides a key to understanding them by distinguishing between 'similar' and 'complementary' activities, although both characteristics are matters of degree, and may be misjudged. He emphasizes the particular difficulties of using either pure market transactions or unitary control to co-ordinate activities which, although 'closely complementary', are also very dissimilar because they rely on very different kinds of knowledge – both 'knowledge that' and 'knowledge how', or capabilities. (We should remember that these differences, continually generated and recombined by the processes identified by Adam Smith, are the prime means of economizing the scarce resource of human cognition, which ensures that rationality is always bounded.) Richardson notes the similarities between Coase's analysis and his own, but also the substantial difference which is marked by his own threefold distinction between consolidation, co-operation and market transactions (Richardson, 1972: 890, 896).

If I have one substantial criticism of Coase's pattern-making as a way of explaining industrial organization, it is that his sharp contrast between firm and market, although rhetorically effective, has tended to impede a proper understanding of the complexities of industrial organization, and especially of its functioning. Richardson has pointed out, and illustrated from his own fourteen-year experience in running a business as a successful Chief Executive of Oxford University Press that the supersession of the price mechanism is not, as Coase (1988: 36) stated, a distinctive feature of management. 'As Voltaire said of God, if prices did not exist, we should have to invent them, and in the design of the internal arrangements of the firm, this is what we in fact do' (Richardson, 1998: 52). We might also question Coase's conception of the firm as a system within which co-ordination is achieved by direction within specified limits. We should remember Chester Barnard's principle that whether a communication carries authority is determined by the recipient (Barnard, 1938: 163), not the originator, and Ménard's (1994) observation that many important communications cross the formal chains of command. Richardson's own prescription that chief executives should give few orders but concentrate on 'creating, monitoring and, when need be, modifying a system of working relationships designed to ensure that each person . . . will further an overarching purpose' (Richardson, 1998: 57) corresponds closely to Barnard's emphasis on the importance of developing mutually compatible ways of thinking within an organization.

The other side of Coase's contrast also requires modification. The transaction costs which are incurred in creating and using market relationships often include a good deal more than the costs of discovering trading partners and the prices which are on offer. For transactions between firms, product specifications and even the design of new products is often far more important, and this may lead to selective relationships between firms which resemble relationships within a well-functioning business. However these important factors should be regarded as amplifications, and to some extent modifications, of Coase's scheme. What is important is that they both essentially result from considering an economic system as an incubator and diffuser of change.

5. A curious incident

In his address to the National Bureau of Economic Research, Ronald Coase recalls Conan Doyle's story 'the adventure of Silver Blaze', in which Sherlock Holmes remarks on 'the curious incident of the dog in the nighttime'; to which the Inspector's response was 'the dog did nothing in the nighttime'. 'That', observed Holmes, 'was the curious incident' (Coase, [1972] 1988: 58). Coase then remarked that the absence of any explanation by economists of how economic activities are divided up among firms is a curious incident of precisely this kind. However, whereas Holmes deduced that when the horse was removed

https://doi.org/10.1017/S1744137414000265 Published online by Cambridge University Press

from his stable the dog did nothing because it perceived no need for action, and used this deduction to solve the mystery, Coase does not explain why economists have been untroubled by the lack of any specific theory of industrial organization. I have tried to sketch an explanation. At Robbins' level of analysis, firms must conform to markets in choosing products as well as prices; and in an era of scepticism, if not hostility, to the working of markets and of business, the natural explanation for the range of activities encompassed by any firm was to be found in monopoly theory, in which firms became obstacles to efficiency.

This curious incident may be explored further by considering the case of Austin Robinson in the 1930s and 1940s. Robinson (1990: 6) has told us that, as an undergraduate studying economics in Cambridge between 1920 and 1922 he had got more from Marshall's *Industry and Trade*, dealing as it did with the problems which most interested him, than from the *Principles*, and that the primary focus among Cambridge economists at that time was not theory but practical research. He added that 'I now find in Marshall ideas that I had not fully appreciated in the 1920s' (Robinson, 1990: 50) but unfortunately does not specify them. Becattini (2006: 615) summarises the 'core of the ideas of the school' as 'the careful, thorough, but also imaginative, study of the phenomena of the organization of production in firms, grouped in industries or clustered in specialized districts'. We should not then be surprised that when Robinson was invited to contribute a volume on monopoly to the Cambridge Economic Handbooks he asked, and was allowed, to write first about competitive industry. (How can we know what is distinctive about monopoly until we understand what competition is like?) The outcome was *The Structure of Competitive Industry* (Robinson, 1931), which was extremely successful: after modest revisions in 1935 it was regularly reprinted without change for many years.

The book is clearly Marshallian in its motivation: Robinson's introduction reflects both Marshall's concern for improving the condition of the people and his emphasis on the importance of the organization and operation of industry in delivering progress – which, as Schumpeter insisted, was far more important than efficient allocation. Although the use of 'optimum' in five out of twelve chapter headings may have encouraged Joan Robinson ([1933] 1969: xiii) to use the optimum size of firms as 'the foundation of my treatment of competitive equilibrium', Austin Robinson did not indulge in 'blackboard economics'. For him 'optimisation' provides a convenient language for explaining the continuing search for improvements by analysis at the margin of current practice and of context-specific knowledge, and for exploring ways of balancing the often conflicting, and changing, requirements of technical, managerial, financial and marketing efficiency: this typically excludes the attainment of optima within any of these subsystems. The course and consequences of these searches differ over time and between firms, not only because of the variety of business environments but also because of the personal characteristics of their managers (Robinson, 1931: 37), thus encouraging 'the tendency to variation' between firms which was

so important as a generator of innovation for Marshall, as it was for Adam Smith. Firms are decision-making systems: neither the set of products nor the methods of production are given, and every firm is different, because it is 'an organism in itself' (Robinson, 1931: 57), which cannot be effectively co-ordinated simply by giving instructions. Robinson includes a rather brief consideration of the relations between each firm and the industry of which it is a member and rather more on the national and international location of industry.

For Robinson, as for Marshall, markets are not natural givens: they are created and modified by the application of the scarce resources of human cognition and capabilities. The industry structure that Austin Robinson analysed is the outcome of both initiative and competitive pressure and also the context for continuing initiative and competitive pressure which changes that structure. This is clearly not perfect competition; but it is equally clearly not a natural precursor of the imperfect competition towards which his wife was working at the time. Twenty years later Joan Robinson (1951: vii–viii) categorized her own work as the consequence of a 'wrong turning'; the path which she should have taken required 'abandoning the static analysis and trying to come to terms with Marshall's theory of development'. In examining the structure of competitive industry as a context for exploration and experiment, with the structure itself subject to continual change, Austin Robinson was already doing in 1931 what Joan Robinson much later regretted not having done at about the same time.

We might therefore imagine a counterfactual history of the 1930s, in which Coase's ideas might have been used. However this seems extremely unlikely. Let me suggest two reasons. First, although it is not difficult to see how they could have been incorporated in Austin Robinson's framework, this would have been incompatible with its function as a straightforward exposition of uncontroversial ideas. The fundamental reason is the powerful attraction of formal modelling, which has now been generally recognized by those interested in the history of interwar economics. Austin Robinson's analysis was, like Marshall's, too close to the realities of business to be readily formalized. Evidence in support of this second reason is supplied by an item of personal record. In my last conversation with Austin Robinson, which occurred during the second conference in Cambridge to mark the centenary of Marshall's *Principles*, he remarked that 'in the 1930s we forgot about marketing'. That is hardly a minor omission; but what is remarkable is that marketing is a prominent theme in *The Structure of Competitive Industry*, where it has a chapter of its own. How could its author forget it so soon, and while his book was being reissued? Was he perhaps overawed by the rise of formalism, in which marketing, having no role in either perfect competition or ideal planning, was (like increasing return) a threat to efficient allocation? I was certainly surprised by his criticism of Philip Andrews' *Manufacturing Business* (Andrews, 1949) for apparently paying too much attention to the evidence of businessmen (Robinson, 1950) – although he had frequently used such evidence in his own book.

Robinson's title seems to match precisely Coase's requirements; indeed it is included in Coase's short list of books that 'were all characterized by an interest in how industry was organized, in all its richness and complexity', which provided his 'view of the subject of industrial organization' (Coase, [1972] 1988: 61). However Coase's ([1972] 1988: 71) proposal for a large-scale systematic study focuses on 'a direct approach ... to discover the characteristics of the groupings of activities within firms' (Coase, [1972] 1988: 73–4), apparently without any theoretical preconceptions.

Lowell Jacobsen (2008) has reviewed and reappraised Robinson's book, with particular – and novel – attention to the relationship between Robinson's and Coase's ideas. There is no evidence that Coase and Robinson either met or corresponded before 'The Nature of the Firm' was published (Jacobsen, 2008: 72, n. 13). However *The Structure of Competitive Industry* receives three citations: they refer to the supply price of factors of production in firms of various size (Coase, [1937] 1988: 43, n. 26), the effects of probable price fluctuations on the cost of organizing (Coase, [1937] 1988: 46, n. 29), and the effects of the size of the technical unit (Coase, [1937] 1988: 46, n. 30). There are also two references to Robinson's (1934) paper on management and the size of the firm, on diminishing returns to management (Coase, [1937] 1988: 44, n. 27) and on the effects of imperfect competition on the size of the firm (Coase, [1937] 1988: 51, n. 44). In the last of these comments Coase rejects Robinson's argument that imperfect competition restricts the size of the firm, because it ignores the possibility of extending the range of products – which had been included in Robinson's book; in the others he uses Robinson's work to support his own argument or to supply detail.

It is obviously difficult to judge how far Coase's reading of Robinson's work shaped his argument, and how far it provided support for ideas which were already formed. (Indeed it is often difficult for an author to make such a judgement in retrospect.) However Jacobsen has done enough to demonstrate that anyone who believes that Coase made a major contribution to our understanding of why and how firms matter should also recognize that Robinson's book, although conceived as a presentation of established ideas, is also an important document – even if its importance for the history of economics derives substantially from the neglect, or outright rejection, of those ideas in the course of creating a 'high theory' which ignored the great theme of the organization of economic development which had been inaugurated by Adam Smith and elaborated by Alfred Marshall.

6. Economic systems and economic theory

The foundational concept of modern economics, precisely stated by Jevons, is the efficient use of scarce resources. Yet much economic reasoning violates this concept by implicitly assuming that human cognition itself is not a scarce resource. In particular, the fashion for rational choice explanations, which

https://doi.org/10.1017/S1744137414000265 Published online by Cambridge University Press

assume that nothing will ever happen which has not already been included in the possibility set, has diverted the attention of many economists from the basic truth, expressed in Vernon Smith's Nobel Prize Lecture, that 'human activity is diffused and dominated by unconscious, autonomic, neuropsychological systems that enable people to function effectively without always calling upon the brain's scarcest resource – attentional and reasoning circuitry' (Smith, 2003: 468). This is the most basic of allocation problems which human face – and which had been recognized and addressed by Alfred Marshall (1994) in 'Ye Machine' before he had turned to economics. (For a detailed exposition, see Raffaelli, 2003). Markets and firms are among the most important responses to this fundamental cognitive problem, because they provide domain-limited contexts for the development of such systems, which we may call institutions, together with devices for identifying the need for deliberate thought. In doing so they exploit one of Adam Smith's basic insights: 'the very different genius which appears to distinguish men of different professions, when grown up to maturity, is not upon many occasions so much the cause, as the effect of the division of labour' (Smith, [1776] 1976b: 28).

Of course, there can be no guarantee that those exploiting the advantages of such institutions will recognize in good time the need to apply their scarce cognitive resources to developing new responses which are appropriate to changed circumstances, or that they will succeed (even with the advice of management consultants.) One motivation for Barnard's analysis of management – and Barnard was clearly aware of the crucial role of automatic responses – seems to have been his observation that very few firms have a long life, coupled with his desire to safeguard the future of his own company. That this no longer exists in anything like its old form, despite the efforts of an unusually perceptive Director of Corporate Planning (who, as previously noted, was a close friend of George Shackle), the regular advice of Peter Drucker, one of the very best management consultants – and also the advice of Ronald Coase, is a warning against underestimating the difficulty. Marshall, who was well aware of the importance of developing effective routines, also warned that a network of proven routines could prevent timely restructuring – to the benefit of outsiders (Marshall, 1919: 135–7). The fundamental concept for understanding the basic possibilities and dangers of any kind of specialization is that of a quasi-decomposable system (Simon, [1962] 1969) – which should have been the prime justification for Herbert Simon's Nobel Prize. Decomposability is essential to both specialization and change; but because the limits to decomposability can never be known any scheme of decomposition, however successful, is always liable to erosion or disintegration.

7. Conclusion

Among the implications of the human cognitive limitations identified by Vernon Smith are 'the severe limitations it imposes on our development of economic

https://doi.org/10.1017/S1744137414000265 Published online by Cambridge University Press

theory' (Smith, 2003: 466). Creating patterns and imposing them on events is what we must do, both within economic systems and in any attempt to understand economic systems; but it is perhaps worth recalling Alfred Marshall's (1920: 355) principle that '[the] tendency to variation is a chief cause of progress'. He added that 'the abler are the undertakers in any trade the greater will this tendency be'. How able is the current generation of economists, compared with their predecessors, such as Ronald Coase?

Acknowledgements

I would like to thank two referees for helpful comments on an earlier version of this paper.

References

Andrews, P. W. S. (1949), *Manufacturing Business*, London: Macmillan.

Barnard, C. (1938), *The Functions of the Executive*, Cambridge MA: Harvard University Press.

Becattini, G. (2006), 'The Marshallian School of Economics', in T. Raffaelli, G. Becattini and M. Dardi (eds.), *The Elgar Companion to Alfred Marshall*, Cheltenham UK and Northampton MA, USA: Edward Elgar, pp. 609–616.

Casson, M. (1982), *The Entrepreneur*, Oxford: Martin Robertson, 2nd edn, 2003, Cheltenham UK and Northampton MA, USA: Edward Elgar.

Coase, R. ([1937] 1988), 'The Nature of the Firm', in *The Firm, the Market, and the Law*, Chicago: University of Chicago Press, pp. 33–56.

Coase, R. ([1960] 1988), 'The Problem of Social Cost', in *The Firm, the Market and the Law*, Chicago: University of Chicago Press, pp. 75–156.

Coase, R. ([1972] 1988), 'Industrial Organization: A Proposal for Research', in *The Firm, the Market, and the Law*, Chicago: University of Chicago Press, pp. 57–74.

Coase, R. (1988), 'The Firm, the Market and the Law', in *The Firm, the Market and the Law*, Chicago: University of Chicago Press, pp. 1–32.

Coase, R. (1991a), 'The Nature of the Firm: Origin', in Oliver E. Williamson and Sidney G. Winter (eds.), *The Nature of the Firm: Origins, Evolution and Development*, New York: Oxford University Press, pp. 34–47.

Coase, R. (1991b), 'The Nature of the Firm: Influence', in Oliver E. Williamson and Sidney G. Winter (eds.), *The Nature of the Firm: Origins, Evolution and Development*, New York: Oxford University Press, pp. 61–74.

Creedy, J. (2013), 'John Richard Hicks 1904–1989', in *Biographical Memoirs of Fellows of the British Academy XII*, Oxford and New York: Oxford University Press for the British Academy, pp. 215–234.

Hayek, F. (1952), *The Sensory Order*, Chicago: University of Chicago Press.

Jacobsen, L. (2008), On Robinson, Coase, and the 'Nature of the Firm', *Journal of the History of Economic Thought*, **30**(1): 65–80.

Kirzner, I. (1973), *Competition and Entrepreneurship*, Chicago and London: University of Chicago Press.

Loasby, B. (2011), 'Uncertainty and Imagination, Illusion and Order', *Cambridge Journal of Economics* **35**(4): 771–784.
Marshall, A. (1919), *Industry and Trade*, London: Macmillan.
Marshall, A. (1920), *Principles of Economics*, 8th edn, London: Macmillan.
Marshall, A. (1994), Ye Machine, *Research in the History of Economic Thought and Methodology: Archival Supplement 4*, Greenwich CT: JAI Press, pp. 116–132.
Meade, J. (1936), *An Introduction to Economic Analysis and Policy*, Oxford: Oxford University Press.
Ménard, C. (1994), 'Organizations as Co-Ordinating Devices', *Metroeconomica*, **45**(4): 224–247.
Menger, C. (1976), *Principles of Economics*, Translated by J. Dingwall and B. Hoselitz, New York and London: New York University Press.
Penrose, E. (1959), *The Theory of the Growth of the Firm*, Oxford: Basil Blackwell. 3rd edn, 1995, Oxford: Oxford University Press.
Raffaelli, T. (2003), *Marshall's Evolutionary Economics*, London and New York: Routledge.
Richardson, G. (1960), *Information and Investment*, Oxford: Oxford University Press.
Richardson, G. (1972), 'The Organization of Industry', *Economic Journal* **82**(327): 883–896.
Richardson, G. (1975), 'Adam Smith on Competition and Increasing Returns', in T. Wilson and A. Skinner (eds.), *Essays on Adam Smith*, Oxford: Clarendon Press, pp. 350–360.
Richardson, G. (1998), 'Some Principles of Economic Organisation', in N. Foss and B. Loasby (eds.), *Economic Organization, Capabilities and Co-ordination: Essays in Honour of G. B. Richardson*, London and New York: Routledge.
Robinson, E. (1931), *The Structure of Competitive Industry*, Cambridge: Cambridge University Press.
Robinson, J. (1933), *The Economics of Imperfect Competition*, London: Macmillan.
Robinson, E. (1934), 'The Problem of Management and the Size of Firms', *Economic Journal* **44**(172): 242–257.
Robinson, E.(1950), 'The Pricing of Manufactured Products', *Economic Journal*, **60**(240): 771–780.
Robinson, J. (1951), *Collected Economic Papers, Vol. 1*, Oxford: Basil Blackwell.
Robinson, E. (1990), 'Prologue', in Rita M. Tulberg (ed.), *Alfred Marshall in Retrospect*, Aldershot UK and Brookfield, VT, USA: Edward Elgar.
Shackle, G. (1967), *The Years of High Theory*, Cambridge: Cambridge University Press.
Shackle, G. (1970), *Expectation, Enterprise and Profit: The Theory of the Firm*, London: Allen and Unwin.
Shackle, G. (1972), *Epistemics and Economics*, Cambridge: Cambridge University Press.
Simon, H. ([1962] 1969), 'The Architecture of Complexity', in *The Sciences of the Artificial*, Cambridge MA and London: MIT Press, pp. 84–118.
Smith, A. ([1759] 1976a), *The Theory of Moral Sentiments*, eds. D. D. Raphael and A. L. Macfie, Oxford: Oxford University Press.
Smith, A. ([1776] 1976b), *An Inquiry into the Nature and Causes of the Wealth of Nations*, eds. R. H. Campbell, A. S. Skinner and W. B. Todd, 2 volumes, Oxford: Oxford University Press.
Smith, A. ([1795] 1980), 'The Principles which Lead and Direct Philosophical Enquiries: Illustrated by the History of Astronomy', in W. P. D. Wightman (ed.), *Essays on Philosophical Subjects*, Oxford: Oxford University Press, pp. 33–105.
Smith, V. (2003), 'Constructivist and Ecological Rationality in Economics', *American Economic Review*, **93**(3): 465–508.

Sraffa, P. (1926), 'The Laws of Return Under Competitive Conditions', *Economic Journal*, **36**(144): 535–550.
Stigler, G. and Boulding, K. (eds.), (1953), *Readings in Price Theory*, London: Allen and Unwin.
Young, A. (1928), 'Increasing Returns and Economic Progress', *Economic Journal* **38**(152): 527–542.