3
ECONOMICS AND EMPLOYMENT RELATIONS

Paul Willman

Abstract
As other chapters in this thematic unit illustrate, the field of employment relations touches on a number of different intellectual traditions, but arguably economics is the largest and most imperialistic discipline to influence the field; indeed, the founding father of personnel economics claims as much (Lazear, 2000). The interchange between discipline and field has a long history, which must be summarised here in the broadest terms, and with a bias towards the most fruitful forms of interchange. This bias leads to a sampling of the discipline in favour of economists who are interested in institutions and specifically those who would accept the proposition that there is a difference between employment contracts and other forms of contract. The second bias is geographical: economics has tended to be more influential for US ER academics than for those elsewhere; one author has even suggested institutional economics is ‘home turf’ for the US ER tradition (Kaufman, 2010, p. 75).

The structure of this chapter is as follows. Section 2 uses a broad historical summary of the discipline’s encounter with the field to try to explain the nature of the emerging relationship. The next three sections use arguably the three most influential concepts imported from economics to ER to illustrate the most fruitful elements of dialogue; these are, in order: incomplete contracts, collective action and voice. I will argue that all have been important to the ER field and have in turn been influential in developing economic ideas about what goes on within firms. The conclusion examines where the relationship might be going, and looks at the most proximate parts of the huge discipline of economics – personnel and organisational economics – to assess the possibility of further dialogue.

Introduction
As other chapters in this thematic unit illustrate, the field of employment relations (ER) touches on a number of different intellectual traditions, but arguably economics is the largest and most imperialistic discipline to influence the field; indeed, the founding father of personnel economics claims as much (Lazear, 2000). The interchange between discipline and field has a long history, which must be summarised here in the broadest terms, and with a bias towards the most fruitful forms of interchange. This bias leads to a sampling of the discipline in favour
of economists who are interested in institutions, and specifically those who would accept the proposition that there is a difference between employment contracts and other forms of contract. The second bias is geographical: economics has tended to be more influential for US ER academics than for those elsewhere; one author has even suggested institutional economics is ‘home turf’ for the US ER tradition (Kaufman, 2010, p. 75).

**A history of dialogue?**

Classical political economy was much concerned with employment relationships. Adam Smith famously discussed both the division of labour and the generation and distribution of economic surpluses; the *Wealth of Nations* also contains succinct descriptions of what later came to be called the ‘principal–agent problem’ in modern economics (Willman, 2014, pp. 37–48). For others, particularly Ricardo, the ‘machinery question’ was a central concern; this ‘question’ concerned the displacement or deskilling of employees by the introduction of new technology, particularly in the textile industries, and what ought to be done about it (Berg, 1980). For Marx, the centrality of the labour theory of value to his intellectual edifice places the employment contract and its dynamics centre stage (see, e.g. Giddens, 1971). All were profoundly influenced by the emergence of employment contracts within firms as a defining property of industrialisation.

It has become something of a cliché to remark on the disappearance of such concerns with the marginalist revolution in economics in the second half of the nineteenth century, so perhaps it is best here to focus on those components of the neoclassical approach that ensure its distancing from the ER field. Two stand out. First, beyond the production function, there is no theory of the firm in neoclassical economics within which to consider the dynamics of employment contracts. Second, employment contracts are not seen as distinctive; they have the same kinds of demand and supply curves as any other form of transaction; whereas the classical political economists saw the employment relationship in hierarchical terms, for the neoclassicist it does not matter much whether employers hire employees or vice versa (Alchian and Demsetz, 1972). It is axiomatic that, as Kaufman (2010, p. 76) puts it, labour is treated as a commodity in competitive markets.

The more general issue – which does not only affect the ER field – is that just at the time when the largest economies in the world were becoming dominated by vertically integrated large firms full of employees, economics goes missing; in Kay’s (2004) terms, economists turned from concern about developments in the economy to an attempt to become physicists. Incidentally, this absence created an intellectual vacuum in which most of the current disciplines of the business school, such as accounting, operations management, marketing and strategy, developed to analyse (and prescribe best practice about) events within firms (Willman, 2014). For ER, with the exception of labour economists studying wage setting, differentials and trade union wage policy – all external to the firm – this absence lasts, as Piore (2011) has noted, from the late nineteenth through to the latter part of the twentieth century.

Although academic economics became estranged from the ER field, it would be an exaggeration to argue that the field of ER developed in isolation from a concern with economics, however. If we look at two Anglo-Saxon countries – the USA and UK – we can find in the work of the core founding figures of the academic field not only a concern with economic matters but the borrowing of economic ideas. For illustration, I examine here the work of John Commons in the USA and the Webbs in the UK. In both, there is clear evidence of the influence of three key economic ideas that have become central to the more modern analysis of ER. First, the conception of an employment contract as incomplete and distinctive, thus subject to conflict of interest; second, a concern with the ubiquity of collective action
by employees; and third, the idea of voice as a mechanism for managing the dynamics of the employment relationship. 3

In the neoclassical approach to labour contracts, wage and employment levels are set by the interaction of supply and demand; labour contracts are by implication complete. The dominant trend in ER is to see the employment contract as involving a hierarchical authority relationship, with an inequality of bargaining power between employer and employee, and as incomplete, that is with the opportunity, indeed necessity, for adaptation to changing circumstances. For the Webbs, this inequality was part of a more general pattern of social inequality that Fabian socialism sought to address, but for both the Webbs and Commons, the key deficiency of the neoclassical view of contracts is the omission of consideration of ownership and property rights. Consider the typical neoclassical production function, combining capital, land and labour. These are undifferentiated factors to combine, but typically (at the time) employers would own land and capital but rent labour, which thus became a more disposable factor of production (see also Kaufman, 2010, p. 88). For Commons in particular, the incompleteness of employment contracts was important because completion (i.e. agreement of precise terms in an ongoing relationship) entailed transaction costs; it followed that some forms of employment contract might be more efficient than others if they entailed lower transaction costs of completion. Although this insight was not fully followed up until the work of Williamson (see below) it pointed to the importance of the contractual governance mechanisms that might minimise transaction costs. For Commons, this pointed to the ‘humanisation’ of the workplace, ensuring the emergence of a win–win game in contract negotiations, as opposed to the then highly prevalent outcome of industrial conflict.

This divergence between field and discipline has proved enduring and it is worth exploring in some detail. To put it very simply, economics for many years had something to say about how wage rates might be struck but very little to say about how employees would subsequently be managed, and this followed from the absence of a theory of the firm. The most influential theory, that of Coase (1937), on which much subsequent work on the economics of the firm is based, has two characteristics relevant for the relationship between field and discipline. First, it generated very little interest within economics at the time; Coase spends much of his article justifying to a Marshallian audience why he is even asking why firms form, and his Nobel prize did not come until 1991. Second, it is a theory of market failure, not a theory of the firm. Specifically, firms arise when certain types of transaction in markets involve such costs of price discovery and enforcement that it becomes more efficient for an ‘entrepreneur’ to use ‘authority’ to direct resources to efficient ends. Strong boundaries emerge where the transaction costs of authority and market as allocative mechanisms are equal. Where ‘entrepreneurs’ come from, and what might be the basis of their ‘authority’, or the mechanisms through which it might be asserted, are not considered; they are assumptions.

This approach had little to say to academics trying to understand what was going on in firms like Ford, GM, DuPont and Sears that were dominating the US economy at the time; looking back, there must have been an awful lot of market failure around, if that were indeed the prime generator of firm formation and growth (Chandler, 1962). Assuming ‘authority’ involves zero or low transaction costs in turn assumes away most of the issues that Commons was trying to raise. As Bromiley (2005, p. 13) much later noted, the difference between a theory that explains why firms exist and one that explains what they do is similar to the difference between a theory of how fish evolved and one explaining how they swim. Coase had much more to say about the former than the latter.

This is an enduring rather than an historical point of interest. The intellectual heirs of the Coasean approach are those who use the principal–agent theory to model how the
entrepreneur’ uses ‘authority’ to control intractable ‘agents’. They regard the large public corporation as an ‘aberration’ and as inherently inferior to markets (Jensen and Meckling, 1976; Jensen, 1989). Those in the ER field who sought, like many other students of modern business, to understand how employment contracts were managed needed to develop their own theories and approaches, either to explain the nature of the contract itself or the system of contractual relationships that constituted the firm (e.g. Doeringer and Piore, 1971; Marsden, 1999; Osterman, 1987). For the modern economist, the Apple or Nike supply chains reveal the efficiencies of renting all of the factors of the production function, particularly for the advantage of shareholders. Many in the ER field retain a concern about what goes on in the factories within these supply chains that would have been familiar to Commons.

The second ER concern to feature in the work of both Commons and the Webbs is collective action by employees, primarily to address the structural power imbalances in the employment contract. Arguably, this was the central contribution of the Webbs to the UK ER field, covering the history of development of trade unions in Britain, their functions in workplace governance and their operation and administration. Commons similarly covered the union role in influencing the operation of the firm, and, most notably in his 1909 analysis of the Philadelphia shoemakers’ union, the dynamics of union development as markets expanded and the division of labour progressed. As Kaufman (2010, pp. 76–78; see also Kaufman and Gall, 2015) has noted, in both cases there was a strong political dimension to this, since there was in the early twentieth century a widespread perception of a ‘labour problem’ in both the UK and USA, the outward sign of which was a high level of industrial conflict in both countries, the agents of which were trade unions. However, again the ER field entered an intellectual vacuum left by economics. High levels of collective action to raise wage rates should not occur for two, rather contradictory, reasons from a neoclassical perspective. First, it should not, given the benefits of free riding by individual employees, be possible, since employees can gain from union activity by enjoying the public goods it produces without paying the subscriptions it needs; this idea was much later formalised by Olson (see below). Second, it should be self-defeating, since raising the supply price of labour should reduce the overall level of labour demand, and thus employment levels would fall. Without collective action, under Say’s law, the outcome will be full employment at equilibrium without government involvement. Needless to say, both the Webbs and Commons were in favour of government intervention in labour markets, as part of Fabian socialism on the one hand and the New Deal on the other.

One may tell another ‘economics catch up story’ here too. As ER writers stepped in to discuss the emergence of new institutions, economics arguably badly needed an efficiency argument. After all, something that is empirically very common in the economy needs an efficiency argument from an economics point of view – otherwise why would it exist? Firms should have a strong motivation to improve efficiency by getting rid of unions and employees should have a strong motivation to quit union membership and free ride. As we will see below, economists were to come up with influential work explaining why both need not happen (though they were better at explaining the first than the second) but there remains the curious fact that, for much of the period during which in advanced economies trade unions were powerful actors, economics struggled to explain why they were there. They did come up with influential theories about the objective functions of union organisations – i.e. the wage/employment trade-offs they pursued (e.g. Dunlop, 1944; Ross, 1948) – so we have in this instance the reverse position from that of the firm discussed above; in the case of trade unions, we have a theory about how fish swim but none about how they got there.

The final economic idea in the ER field we discuss that traces its lineage back this far is that of voice, although admittedly this more modern term is not used by either the Webbs
Economics and employment relations

or Commons. The central concern of both was the imbalance in authority and power in the employment relationship. The authority issue for the Webbs was that embedded in the law covering employment contracts in the UK, which was explicitly inherited from ‘master and servant’ acts going back several centuries. The core of the issue was the employer’s generic right to give instructions and the employee’s more limited right to disobey only if the instruction was unlawful or unreasonable (Wedderburn, 1971). The power issue was the imbalance of bargaining power in the employment relationship between the employer and the individual employee. In their treatise on industrial democracy, the Webbs were concerned to establish checks and balances on employer power and authority both by the introduction of counterweights to employer power (primarily the role of unions) and constraints on employer authority (primarily legislative proposals). Although the Webbs clearly thought such measures would be socially beneficial, they do not develop the efficiency arguments. By contrast, Commons did. ‘Constitutional’ government of contractual relations in industry would, he argued, lead to lower transaction costs in filling the ‘silences’ in employment contracts in firms and thus to greater efficiency. Where employees felt engagement in the terms of their contractual relationship, more co-operative relations were likely. In Commons, we find one of the earliest recognitions of the mantra that the management of people can generate competitive advantage.

Let us summarise the argument so far. Classical political economy was deeply concerned with ER because it saw the emergence of employment contracting and firms as a crucial part of the industrial revolution. Neoclassical economics shed this concern in an attempt to produce generalisable mathematical propositions which, inter alia, failed to recognise the idiosyncratic nature of employment contracts. Other academics and policy-makers – we have used the Webbs and Commons as exemplars – stepped into the vacuum economics left in order to analyse what was going on within the firms and institutions that neoclassical economics excised from its agenda as noise. They developed ways of analysing three key areas – employment contracts, collective action and contractual governance arrangements – which have remained influential within the ER field to date.

The final part of this argument concerns the return of the prodigal son. Although economics was not a source of core concepts for those concerned with employment contracts in the period we have discussed, it has become so. In the next three sections, we will argue that developments in institutional economics have become extremely influential in understanding how these three areas of activity central to the ER field operate. Simon and Williamson have developed the analysis of employment contracts, Hirschman and Olsen have developed the analysis of collective action, and Freeman and Medoff have contributed to the modern understanding of voice. We look at each in turn.

**Authority and contract**

Simon (1951) is one of the earliest and certainly the most influential to attempt to capture the distinctiveness of employment contracts from an economic perspective. Simon notes that most labour is (in 1951 in the USA) delivered under employment contracts and that these differ from sales contracts, which are the standard stuff of price theory. In the latter, the commodity to be exchanged and the price are completely specified, and the seller of the commodity does not care what happens to it after sale. By contrast, an employment contract exists when the employee ‘agrees to accept the authority’ of the employer (1951, p. 294) in return for a given wage. Simon addresses the two obvious questions: first, why would the employee accept the employer’s authority and second, why would the employer want more than a sales contract?
The answer to the first question is simple: the employee is paid a premium for this concession; it is implicit in the argument that the more the employee cedes in terms of the range of this authority, the higher the wage premium will be. Simon specifies a range \((x_1 \ldots x_n)\) of decision items across which the employee will in effect delegate decision-making to the employer in return for a premium in the employment contract over the sales contract price. The employer, by assumption, will in turn wish to have an employment contract as a way of purchasing discretion in decision-making. If they cannot predict which \((x_1 \ldots x_n)\) will be needed at which time, then paying a premium for the right to choose may make sense.

This approach has influenced a range of subsequent developments in the analysis of employment contracts in economics, both for institutional economists such as Williamson and also for those who developed principal–agent theory into a formal model of the employment relationship (Jensen and Meckling, 1976), so it is worth making a number of general points. First, employment contracts are ubiquitous when Simon is writing, so they must have efficiency properties, otherwise all employment would be under sales contracts. These efficiency properties fundamentally lie in the absence of transaction costs and the management of uncertainty. Simon does not use the term transaction costs, but he is aware that the employment contract cannot be both continuously renegotiated and efficient. Uncertainty for the employer about the future demand for labour services carries a premium wage, and the greater the uncertainty, the greater the premium. So it is but a short step to the argument that authority is efficient. Its origins thus become unproblematic. Second, actors are presumed to be rational and would only enter employment contracts where it was in their own interests. Simon explicitly refers to the idea that an employment contract can be seen as a repeated game (1951, p. 294). They may be boundedly rational, being unable to predict how the future \((x_1 \ldots x_n)\) will develop, but non-rational elements in employment contracts are not part of the model. The construction of an employment contract is thus from the employer’s point of view similar to that of taking out an option on a trade; since they do not know when labour service \(x_j\) will be required, they purchase a call option. The employee takes the present value of this option as their wage.

This approach was developed by Simon’s pupil, Williamson, who noted that some employment contracts look more like sales contracts than others. Before outlining his approach to employment contracts, it is important to be clear about the assumptions and core concept in his approach. Williamson is primarily concerned with two related issues: the Coasean problem of the comparative efficiency of firms (‘hierarchies’ in his terms) on the one hand and markets on the other, and the efficiency properties of different types of hierarchy (Williamson, 1975, 1985). The step forward from Coase is precisely in his concern to theorise the dynamics of hierarchy; hierarchies may themselves differ in efficiency depending on the choice and mix of contractual forms with different levels of transaction costs. He assumes that firms will arise where assets are specific to a transaction – i.e. they have lower values outside the firm than inside (this follows from Coase).

Two further key assumptions about economic actors are important. He assumes individuals are boundedly rational and that complete contracts are difficult to write (this too follows from Simon). More controversially, he extends the conventional economic assumption that individuals pursue self-interest to the idea that they are ‘opportunistic’ in that they pursue ‘self-interest seeking with guile’ (1985, p. 47).

The scope of his work goes beyond employment contracts, but here we restrict our focus to employment. He distinguishes three broad contract types. Sequential spot contracting is a market-type transaction where employer and employee sustain no long-term commitments; they simply exchange cash for services as the need arises. A simple example might be that of a casually employed gardener or window cleaner paid by the householder as need arises.
A more modern example would be the use of such contracts in the ‘gig economy’, typified by app-driven employment in firms such as Uber (Habans, 2016). However, more complex institutional structures may emerge to govern sets of spot transactions. Longshore work in a variety of countries based on casual labour or, historically, the printing of newspapers appear to be institutionalised forms of spot contracting (Edwards, 1990; Sisson, 1975; Willman, 1986). Day hiring of labour in industries such as construction would be another example, as would the use of zero hours contracts.

Williamson identifies the strengths and weaknesses of this contractual type. From the employer perspective it has the advantage that labour does not become a fixed cost, but the disadvantage is that the employer cannot easily develop firm-specific skills. For the employee it allows choice of employment opportunities but has the disadvantage that all of the contractual risk falls on them. It is clear that the balance of advantage depends crucially on the supply/demand relationship. Spot contracts work well for employees where labour is, or can be made, scarce and a key strategy of unions (or other forms of collective action) in all of the industries above is to restrict the supply of labour and to maximise the number of bargaining opportunities.

Williamson’s second type of employment contract is the contingent claims contract. In this arrangement the parties try to write a comprehensive contract itemising every employee task and setting a price for it. Again, there are historical examples. The making of shoes in the UK and USA was for many years based on extensive price lists for components such as soles and heels in which the cost of labour was included (Commons, 1909; Willman, 1986). The advantage for both employer and employee is reduction in uncertainty, and the problem for both is bounded rationality. Where events arise not covered in the comprehensive contract the haggling and thus transaction costs will emerge.

Williamson thus falls back on Simon’s original idea (see also Simon 1991) of an authority relationship as having the most robust efficiency properties, and the example he relies heavily upon is Doeringer and Piore’s (1971) ‘enterprise market’ – i.e. an internal labour market. Williamson (1975) describes it as an authority relationship with many contingencies spelt out. He emphasises several characteristics with efficiency properties: internal promotion ladders encourage the development of firm-specific skills; pay rates attached to jobs and a regularised promotion system lower the requirement to monitor employees intensively; job security encourages employee co-operation; and agreed procedures reduce haggling. In this contractual form, labour becomes a fixed cost and the employer bears much of the contractual risk (upside and downside) in return for a reduction in transaction costs.

One might find it ironic that such institutionalised forms of labour management arguably began to disappear in the USA just about the time Williamson wrote about their efficiency properties. However, there are two features of his analysis of ER that have enduring interest, not least because they embody major differences between economic approaches to employment relationships and those grounded in sociology or politics. They concern the outcome variables of efficient employment contracts, i.e. the avoidance of opportunism and the generation of consummate co-operation.

Opportunism – ‘self-interest seeking with guile’ – is an assumption about individual behaviour in contract negotiations, but in the analysis of employment contracts it emerges as a property exclusively of seller (i.e. employee) behaviour. The institutions of contractual governance are efficient to the extent that they control employee opportunistic behaviour such as deliberate underperformance or fraud. Employers (buyers) are assumed to be immune to opportunism (Willman, 1982). Consummate co-operation is defined as ‘an affirmative job attitude – to include the use of judgement, filling gaps and taking initiative’, contrasted with perfunctory
co-operation, which is ‘job co-operation of a minimally acceptable sort’ (Williamson, 1975, p. 27). It is key to the generation of efficiency gains and firms that encourage it will out-compete those that do not. Again, it is a property of employees, not employers.

Williamson anticipates much of the later agenda about the relationship between human resource practices and competitiveness, in which a variety of terms have been used to describe desirable employee outcomes – most recently ‘engagement’. However, his approach is also intellectually very similar to those embedded in more recent economic attempts to model employment contracts in economics, notably principal–agent theory (Jensen and Meckling, 1976), in three respects, all of which tend to make economic models of the employment contract unpalatable to many in the ER field. They are:

1. Authority of the employer is an assumption, not a variable. The Coaseian roots of the idea of authority in hierarchy as a source of efficiency means that its legitimacy is not questioned and its objectives are unexamined. Challenges to employer authority show up as transaction costs and thus efficiency losses. Most ER scholars regard power and authority as central to the analysis of employment contracts, and thus see economic approaches that sideline them as incomplete at best and biased towards the employer at worst (e.g. Edwards, 1990; Nolan 2012).

2. Efficiency is defined in terms of the appropriation of all economic rent by the employer. Appropriation of such rents by the employee is opportunism, and opportunism is seen as a property only of seller (employee) not buyer (employer) behaviour (Willman, 1982). The subsequent use of such rents is exogenous to the model of the employment contract.

3. The central problem of employment contracts is thus employee behaviour, specifically the avoidance of opportunism and perfunctory co-operation, and elimination of these undesirable outcomes, is a function of contract design.

It should be noted that developments in principal–agent theory share these features. Principals are risk-neutral and rational; agents are risk-averse and prone to generating information asymmetries. Those selecting agents must beware moral hazard and adverse selection. Management of agents involves a mix of monitoring and incentives and the design of this mix generates the desirable employee (agent) behaviours. As I have argued elsewhere, principal–agent theory is essentially a Taylorist approach to ER based on close monitoring and financial rewards (Willman, 2014, pp. 37–48).

The problem of collective action

Any theory in economics that stresses the difficulty of collective action is probably best regarded by those in employment relations as a constraint on theorising rather than a contribution to a field in which employee collective action is widespread. Olson’s (1971) approach, mentioned above, must nonetheless feature here since it has provoked substantial debate within the ER field (Crouch, 1982; Kelly, 1998). However, much of this section will discuss another approach – that of Hirschman (1970) – which focuses in a more conventional economic way on the efficiency properties of a widely observed social phenomenon. Olson’s approach does not have much to say about workplace representation, while Hirschman’s explains much about the market for collective action; it also relates to the discussion of voice in Section 5 below.

Olson’s approach relies on the one hand on the idea of calculative rationality and on the other hand on the distinction between public and private goods. By calculative rationality Olson primarily means that people will engage in collective action only where the benefits
of doing so exceed the costs. The costs may be monetary (e.g. union subscriptions) and non-monetary (e.g. time involved). The central problem here is that for Olson, where collective action only produces public goods, free riding is more profitable than collective action. This is because a public good is one whose benefits are available to all, regardless of whether they have incurred the associated costs. Therefore, for Olson, private goods – restricted to those who have incurred costs – are essential if we are to avoid having a situation where everyone would benefit from the provision of public goods but there is no individual incentive for co-operation. He uses the term selective incentive to refer to private goods, and the term special condition to refer to forms of coercion or compulsion to compel individuals to engage in collective action.

Let us put all this into the language of ER. In this view, unions cannot only engage in collective bargaining, since typically the benefits of collective bargaining are a public good; non-members usually get the union-negotiated wage increase. They must provide private goods, typically individual benefits such as insurance or representation for individuals in disciplinary and grievance procedures. Where possible they will seek to make union membership compulsory by collective agreement, or encourage employers to advocate or facilitate union membership (for example by offering to deduct union subscriptions from salary). Much evidence indicates that unions, in the UK at least, do all of these things (e.g. Pencavel, 1971; Willman et al., 1993). Olson’s analysis indicates how unions try to solve first- and second-order collective action problems; the first-order problem is how to overcome free riding to encourage membership growth, and the second is how to run collective action organisations so that the benefits to members are greater than the costs incurred (Willman, 2004; Willman et al., 2016).

There are two major problems with the Olson approach that have attracted the attention of ER researchers, however; the first is logical, the second empirical. As Elster (1989, p. 64) initially noted, the logical problem is that the union has to already exist in order to apply the selective incentives and special conditions that cause it to exist. The empirical objection (see Crouch, 1982; Kelly, 1998) is that there appears to be no good fit between historically observed bouts of union membership growth and the existence of special conditions. Partly as a result of these criticisms, the ER field has been influenced rather more by an approach to collective action that emphasises its efficiency consequences.

Hirschman (1970) was not initially setting out to solve an ER problem, and his approach is largely couched in terms of consumers rather than employees. Assume an organisation produces a good or service for consumers. Assume further that these customers experience a decline in quality of the good or service in question. The standard economic approach would say that consumers would then switch to the outputs of another organisation, until managers of the first organisation remedy the problem, when they might switch back. This is termed exit by Hirschman and he argues that it may be inefficient for all parties. For the organisation, exit is informationally inefficient; you know you are losing consumers, but it may take a long time to find out why and even longer to get them back. For the consumers, there may be substantial switching costs; in his words, consumers have loyalty – this is not irrational or affective, it simply refers to the costs of exit versus the alternative, which he terms voice. Voice involves consumers complaining directly to management about the quality decline, and doing so collectively. Managers get information and the consumer avoids the costs of switching. The reader may wish to think about what happens when an Apple product has a defect in order to assess how useful this approach is in understanding product market behaviour; switching away from your new iPhone means switching other products and services as well.

The approach has been applied successfully in the ER field, however, notably by Freeman and Medoff (1984) to explain aspects of unionisation and, particularly, voice. We discuss this in detail in the next section, but here we focus on the fact that Hirschman has, without particularly
Willman

focusing on it, solved a number of the ‘Olson’ problems about collective action, and we close this section by summarising them.

Olson asserts that setting up collective action organisations is difficult and costly. Hirschman deals with this in three broad ways. First, he introduces the idea of switching costs. Where sunk costs of collective action are high, it may nonetheless be the case that switching costs are higher. To fall back on Williamson’s language, the assets of both employer and employee may be so specific to the employment contract that both can bear the costs of employee collective action in order to sustain the contract. This leads to the second point, that the calculation of the costs and benefits of collective action is not put simply from the point of view of the individual considering union membership. The avoidance of costly employee exit may be worth investment by employers, shareholders and even consumers in voice mechanisms that would prevent it. Third, he sidesteps the free rider problem central to Olson’s analysis simply by allowing that individuals have different utility functions in considering collective action. Specifically, he assumes that there exists a subset of consumers (and by extension, employees) for whom the benefits of collective action are not the financial returns minus the costs but the sum of the two. These individuals are the activists who promote collective action for returns calculated in terms of both extrinsic and intrinsic rewards.

Hirschman’s approach has been influential in the ER field but it is not without its critics. Several point to the difference between the consumer market roots of the approach and the hierarchical realities of employment contracts (e.g. Allen, 2014), suggesting that there are such fundamental differences that the model is inapplicable. Others have suggested that the exit, voice and loyalty categories are too simple, adding others, notably the idea of employee deliberate ‘neglect’ as a response to any deteriorating employment situation (e.g. Farrell, 1983). But the main issues concern the scope of the term ‘voice’, which has a restrictive and remedial efficiency function for Hirschman, but a much broader set of interpretations in the ER field. We turn to it next.

Voice

One of the most influential books of the past 30 years in employment relations is by two labour economists, Freeman and Medoff (1984). It is remarkable in a number of ways. First, it is derivative, in that it simply applies Hirschman’s insights to the labour market. Second, it uses only a US data set but makes veridical claims about general relationships. Third, it assumes, without explicit justification, that unions are the most effective form of collective voice. However, it is highly cited and highly influential (Blanchflower and Bryson, 2004) and has defined the economic approach to voice within ER. In the book’s title, the authors ask the question What Do Unions Do? However, they are not concerned with union activity itself but with its impact, on wages and the general performance of the economy. So the focus is not on unions per se, but on the impact of their presence in the firm on two key dependent variables, wages and productivity.

In their transaction costs approach, the net effect of unionisation is an empirical question that essentially consists of netting off the benefits of unionisation from the known costs. The known costs are monopoly wage gains and restrictive work rules. The key benefit is voice, which has simple Hirschman qualities of reduced exit and reduced information asymmetry. Let us look at each in turn. The conventional economic assumption underlying monopoly wage gains is that they should lead to reduced employment levels through the incentive to substitute capital for labour. The higher rates should also attract better quality labour (unless everyone else is doing the same). The approach does not deal directly with variation between firms in the extent to which capital might effectively displace labour nor the variable cost of capital. Unions
naturally respond to this substitution effect either by opposing technical changes or by retaining prior employment levels in the face of technical change (Willman, 1986). The model contains no variables that might estimate the ability of unions to do so; supply/demand considerations do not really enter into it. The voice variable is in fact an interaction. It depends on the ability of employees to organise voice and the receptiveness of management to the information voice produces. For both of these reasons, it is probably not best conceptualised purely in terms of union activity. Since its effectiveness depends on management reception of its outputs, any form of voice that does not have to offset the imposition of restrictive work rules is likely to be more efficient in Hirschman terms.

However, from a field perspective, the approach has undoubted benefits. The list below is not intended to be exhaustive:

1. It stresses that the impact of unions on the dependent variable, productivity, is an empirical question – not an assumption. The unit of analysis for consideration of this question is the firm or the establishment.
2. The balance between pursuit of wage gains and restrictive work rules puts emphasis on the objective function of unions by focusing on the wage/employment trade off. Unions may seek to maximise the level of employment, or seek higher wages for a smaller number of employee members (Dunlop, 1944; Ross, 1948).
3. The balance between voice benefits and other costs depends on the extent of managerial responsiveness to voice. This is thus a more sophisticated version of the ‘you get the union you deserve’ argument, placing emphasis squarely on the effective management of the trade-offs in the model.
4. By extension, it defines the set of firms in which the net impact of unions on productivity can be positive. Where there are few opportunities for capital investment, where unions can impose restrictive work rules and where management ‘quality’ (as they define it) is poor, the impact of unions on the dependent variable will be negative (Willman, 2014, pp. 55–56).

What this approach does not do is examine the different mechanisms of voice in terms of their effects on productivity. This is more fully dealt with elsewhere in Chapter 16. The key point to make here is that because Freeman and Medoff assume that effective collective voice equates to union activity, two important issues are under-investigated. The first is the employer interest in voice promotion; there is substantial UK evidence that as union coverage declined employers invested substantially in other voice mechanisms (Budd et al., 2010; Gomez et al., 2010). The second is that there is evidence from the same data sets that employer-initiated voice mechanisms have efficiency advantages over unionised ones (Bryson et al., 2013).

The significance of employer investment in voice for an economist is that it emphasises the efficiency properties of voice for organisational performance. Many ER academics might look at employer-sponsored voice in terms of power and legitimacy. It might be seen as offering lower levels of representation and protection for employees than independent trade unionism. The Hirschman approach, as we have seen, sees its efficiency properties in terms of its ability to address an information asymmetry problem that adversely affects firm performance. The significance of any performance advantages of non-union voice is that it simply reflects the trade-offs in a make or buy decision. Firms may make voice mechanisms themselves in order to reap their benefits (non-union voice) or they may buy them (from a union).

A final point to make is that the idea of ‘voice’ within the ER field has developed beyond the economic view of it as an efficient alternative to exit as a reducer of information asymmetry. In the field, employee voice is studied for its effects *inter alia* on employee affective states such as
job satisfaction and organisational commitment, and thus engages with debates within organisational psychology, and it is also related to issues concerning statutory rights within the workplace, and thus engages with the concerns of legal scholars. The impact of voice mechanisms on managerial behaviour represents only a fragment of these broader debates (see Wilkinson et al., 2014).

Conclusion

This chapter has tried to show the interplay between the field of employment relations and the discipline of economics. Neither is homogeneous; we have tried to proceed by showing how the two address problems that are seen as interesting from both perspectives, which has naturally led to a rather selective view of the concerns of both discipline and field. In ER, there are many who would see economics as secondary to sociology, politics and law, and in economics there are still those in denial about the existence of firms. So, in conclusion, we focus on differences between those in the field who think economics is relevant and those in economics who think the employment contract is distinctive.

The key distinction is, in fact, quite fundamental and conceptual. Even those in economics who acknowledge the distinctiveness of employment contracts feel that efficiency drives organisational outcomes and they do not focus on power. There is of course not space here to discuss how power might be conceptualised, but there are no mainstream economic models that deal with this, however defined. In Simon, Williamson, and agency theory the Coaseian assumption that the exercise of what they term authority rests with the employer is an unexamined assumption, not an empirically investigated variable, except in the very limited sense that it generates information asymmetry problems that need to be resolved.

This impacts on how economists deal with the silences in the employment contract, and again the base assumption is that the most efficient conflict resolution mechanism should prevail. Collective action and voice work to the extent that they improve firm performance. If they do not or, more realistically, more parsimonious collective action or voice mechanisms emerge, they will outcompete existing rivals. At the risk of labouring the point, it is worth emphasising the institutional agnosticism of economic approaches. Effective voice is whatever generates the best efficiency (productivity) outcome, normally measured as some return to the firm, rather than employees. This leads to a further point, which is that economic models make employer behaviour endemically problematic, but employee behaviour operates by assumption.

In closure, we focus on likely future relations between discipline and field. As readers will see from the dating of many economic references in this chapter, the high point of institutional economics appears to have passed and developments in personnel and organisational economics that focus on the same subject matter as ER operate with very different methods and assumptions. Their tendency to regard an organisation as an inferior substitute to a market, existing only where a market cannot work, tends to two inclinations. The first is to view the organisation as a nexus of constrained contracts that should operate as much like a market as possible. The second is to work out how the constraints on market operation can be removed to make the organisation dissolve into a market.

The growing field of personnel economics in particular is one that analyses some of the core activities of ER, but uses economic and mathematical approaches to model outcomes. Lazear and Oyer (2013, pp. 479–481) for example, identify the scope of personnel economics as “focused on five aspects of employment relationships – incentives, matching firms to workers, compensation, skill development and the organisation of work”. This set of issues is clearly part of the ER field, but they also identify its objective as “to describe how human resource practices
can best address employers’ goals”. Collective action, power and authority relationships, conflict and bargaining take second stage where one seeks to optimise firm performance without a more pluralist analysis of the interests of other stakeholders. This pluralism has been a feature of the ER field since its origins, and it has only effectively engaged with economists who share it. An optimist might anticipate a return by economists to the classical concerns of political economy, in which employment contracts and their governance were central to an understanding of economic performance. A pessimist might point to the current contents of economic journals.

Notes

1 We use the two terms consistently throughout. ‘Discipline’ refers to economics; ‘field’ refers to employment relations. These terms are simply categorical, with no intended implications about the relative standings.

2 We use the two terms consistently throughout. ‘Discipline’ refers to economics; ‘field’ refers to employment relations. These terms are simply categorical, with no intended implications about the relative standings.

3 The references for the following paragraphs on the Webbs and Commons are listed in the reference list.

References


Economics and employment relations


