2 Employment contracts

Ann-Sophie Vandenberghe*

1 Introduction
Individual employment contracts are contracts, and the primary focus of employment law concerns the contractual relationship of employment. Like other contracts, the employment contract is a consensual relationship between two parties involving an exchange – in this instance, work in return for pay. The standard rules for the formation of legally binding contracts apply to employment contracts. Breach of the contract gives the injured party a right to claim a legal remedy. The standard rules of general contract law intend to solve problems that are universal to all contracts. The law and economic literature on general contract law explains how standard contract rules provide optimal incentives for performing, breaching and relying and correct for pre-contractual information failures.

Why then do we need specialized rules for employment contracts? We need to focus on those features of the employment contract that render the general law of contract unsuitable for handling all disputes that may arise in connection with employment contracts (Collins 2003, p. 6). The traditional legal justification for and explanation of employment law is to protect the employee as the weaker party to the employment relation. Because employees are in an inferior bargaining position compared to the employer, the justification runs, they need protection through employment rules in their favor as an inequality-compensating mechanism. While this may be true for particular workers at particular moments in time, there is no reason to accept as a paradigm of employment law in modern times that employees are *ex ante* (before the contract is made) or *ex post* (after the contract has been made) in a weaker bargaining position. This is not to say that there is no need for a separate body of law to govern employment contracts. Indeed, as will be shown in this chapter, there are good economic reasons for specific rules for employment contracts. After all, many employment relations do have features that are absent in standard contracts so that an efficiency argument can be made for a specific role for the legal and court system with regard to employment contracts.

* I would like to thank Professor Gerrit De Geest for sharing his inspiring and thought-provoking views on the economic analysis of employment contracts.
2 Typical features of employment contracts

Before one turns to a discussion of the typical features of employment contracts, it is necessary to understand why employment parties would want a contract in the first place. A contract is an exchange of goods and services. However, an exchange does not necessarily require a contract. Why contract rather than choosing a simple wage-for-service exchange? A contract is a legal instrument that puts legal pressure on the actions that the contract parties have agreed to take at various times, generally as a function of the conditions that hold (Shavell 2004). Parties want their contracts to be enforced by the courts to avoid the danger of opportunistic behavior. The danger of opportunism arises when performance of contractual obligations is non-simultaneous (Posner 1992, p. 89). For then, in the absence of legal enforcement (and assuming purely self-interested behavior, no repeat play, no reputation sanctions and no taste for fairness), the last performing party has an incentive to opportunistically withhold or change his performance obligation.

The problem of opportunistic behavior can be solved by drafting a contract that specifies the actions that parties are supposed to take at various times and that makes the non-performing party subject to legal sanctions. Employment parties frequently make investments that are specific to the relationship. Employees invest heavily as they pursue a career with a single employer. They invest in the acquisition of skills and knowledge that are more useful to their own employer than they would be elsewhere – what economists call firm-specific human capital (Becker 1975). Employers too invest in the relationship. They incur search and hiring costs and also pay for the creation of the firm-specific skills of their employees. When relationship-specific investments are made, performance is always non-simultaneous, because these investments are made before the exchange is complete (De Geest forthcoming). Hence, with relationship-specific investments, the parties may want a legally enforceable contract to avoid opportunistic behavior. However, due to the typical features of the employment relationship – to be outlined in the following paragraph – it may not be possible to prevent opportunistic behavior through the use of a legally enforceable, fully contingent contract. Hence the parties must seek specially adapted contractual devices that result in ‘relational contracts’.

A Employment as a relational contract

Goetz and Scott (1981, p. 1091) define a contract as relational when the parties are incapable of reducing important terms of the arrangement to well-defined obligations. Why are employment parties not able to draft a fully contingent complete contract? Klein (1980, p. 356) gives two main reasons for the incompleteness of contracts: ‘First, uncertainty implies
the existence of a large number of possible contingencies and it may be very costly to know and specify in advance responses to all of these possibilities. Second, particular contractual performance, such as the level of energy an employee devotes to a complex task, may be very costly to measure. Therefore contractual breach may often be difficult to prove to the satisfaction of a third party enforcer such as a court. Given the presence of incomplete contractual arrangements, Klein (1980, p. 356) asserts that ‘wealth-maximizing transactors have the ability and often the incentive to renege on the transaction by holding up the other party, in the sense of taking advantage of unspecifed or unenforceable elements of the contractual relationship’.

B Firm-specific human capital
Economists use the term ‘human capital’ in the sense, strictly analogous to physical capital, of an asset that yields earnings over time rather than immediately. Hence, human capital is used in the sense of earning capacity. Formal education and on-the-job training are examples of activities that create human capital. A rational person, whether employer or employee, would not incur the costs, direct or indirect, of creating or acquiring human capital unless they were offset by higher productivity or, in the case of the employee, by higher earnings (Posner 1995). Economists distinguish between two types of human capital: general human capital and firm-specific human capital. It is the latter type that concerns us here.

Much of the human capital that on-the-job training creates, unlike that created by formal education, is usable only in the particular employer’s employ (‘specific human capital’). The costs of creation of specific human capital consist of the direct costs of training and the indirect costs of foregone earnings during the period when the employee is learning and not working. Who is to pay for the costs of training? Suppose that the employee pays for all the training costs; he or she finances the direct training costs and does not receive a wage during the training period. The employee will only be willing to make this investment if the employer promises that his or her future earnings will contain a component that is implicitly a repayment (with interest) of the employee’s investment in human capital. Due to a lack of information ex ante, it is not possible to ‘draft’ exact wages for the future. Therefore, wages will have to be negotiated ex post, but in these negotiations the employer is in the strongest position. He may propose a wage that is just above the market wage, but not high enough to pay back the employee’s investments. The employee may find this unfair, but his only alternative is to leave the firm, in which case he will get the normal market wage, losing his entire firm-specific investment. The employee will be reluctant to pay for such training because, if the capital it creates is
specific, he cannot use the threat to quit his current job and take his human capital elsewhere to extract a higher wage. Would this problem be solved by letting the employer finance the full investment? It wouldn’t. That solution would create similar problems. Now the employee would be able to hold-up the employer by demanding a higher than market wage. Such a demand is unfair because the employee did not contribute to the investment, but the employer’s only alternative is to hire another employee, in which case he has to make the same investment again.

According to Gary Becker (1975), the incentive to make opportunistic threats is minimized if both employer and employee contribute to the cost of creating firm-specific human capital, for then each has something to lose if the other terminates the employment. At the same time, parties have less to lose when the other threatens to terminate the employment. At the \textit{ex post} negotiation stage, parties will have an equal negotiation position. The conditions that apply at the \textit{ex post} negotiation stage are those of a \textit{bilateral monopoly}. The solution of sharing the investments costs is only an imperfect solution because the \textit{ex post} negotiations may still fail due to strategic behavior. Each party may waste resources to appropriate the largest share of the surplus. However, when it works, investment by a present or a future worker in his human capital creates a steep age-earnings profile. During the investment period, earnings are low or negative, reflecting the fact that the employee pays part of his firm-specific training. But after the initial training period, he will earn more than what he could earn elsewhere, reflecting the fact that earnings contain a component that is implicitly a repayment (with interest) of his earlier investment. Wages have to be high during the recoupment period in order to make the investment worthwhile. The employer can afford to pay the trained worker well, because his marginal product will be high because he has so much human capital and the employer has an inducement to do so in order to prevent him from quitting, which would wipe out the firm’s investment in human capital.

According to De Geest (forthcoming), the employment parties could also choose to have their incomplete, relational contract governed in another way: third-party governance. Under third-party governance, a judge or an arbitrator decides on wages and other gaps in the contract \textit{ex post} when more information is available (De Geest forthcoming). This solution, he asserts, requires that two conditions are fulfilled. First, the third party needs to have enough information. This condition may be difficult to fulfill when the nature of the exchange between worker and firm involves surplus that is specific to the transaction as opposed to surplus that is specific to the market as a whole. Second, the third party needs to be fully neutral.

One could also wonder why employment parties would be prepared to invest in relationship-specific investments when this makes them
potentially vulnerable to exploitation by the other party. Sometimes relationship-specific investments are required in order to realize the full gains from exchange. But perhaps the full gains from cooperation could also be realized when the employer switches to production technologies, methods and procedures which require investments in *general* human capital.

General human capital consists of skills and knowledge that the worker can employ elsewhere in the economy. De Geest (forthcoming) gives an example: ‘If each firm uses its own bookkeeping system, software, and working method, new employees would first need a firm-specific training. If all firms use more or less the same bookkeeping system, the same software and working methods, workers can easily switch between firms’. This, he contends, might require legal intervention in setting technical standards, because markets sometimes fail to produce the optimal degree of standardization.

Who will pay for investment in general human capital? The employee himself pays for the employer’s investment in the employee’s general human capital by accepting a lower wage. The employer would have no protection against the employee using that capital to obtain a higher wage from another employer, and therefore he will not pay for it. On the other hand, when the employee pays for investment in general human capital, he is not vulnerable to opportunistic hold-up threats made by the employer because he can take his general human capital elsewhere. In those instances where the employer does pay for investment in general human capital, his investment can be protected through the use of contractual safeguards, such as the covenant not to compete.

Reducing the amount of relationship-specific investments made, or substituting general human capital for specific human capital, does not solve all problems associated with the employment contract. As will be discussed in the next paragraph, there is still scope for opportunistic behavior when the performance standards cannot be specified in advance and information about performance is asymmetrically distributed during the performance of the employment relationship.

C  Agency problems
Rosen (1984, p. 986) gives a nice description of the employment contract:

. . . an employment contract is a very curious creature, since the nature of the exchange is not made very explicit in most instances. On one hand, employment is an authoritarian relationship in which the employee undertakes an obligation to follow certain orders and commands of the employer. On the other hand, it contains important elements of delegation in which the employee is given latitude within broad limits to behave in the interests of the firm. Some elements of the contract are express or clearly implied and are actionable for breach ( . . . ).
But most terms of the contract (…) are left unspecified, precisely because it is too costly to write them down and to verify that they have been performed. These therefore represent an implicit understanding about which there may be significant scope for disagreements through asymmetrical information.

Employers draft contracts for employees to sign, but expect far more than its terms. The employment contract creates a legally enforceable exchange of work in return for pay. Whereas the obligation to pay is a well-defined legally enforceable obligation, the obligation to perform work is by its nature not clearly specified and can be performed in several ways given the minimum requirement of showing up for work during working hours. The traditional employment contract vested a discretionary power in employers to direct the workforce to its most profitable use. This arrangement becomes inefficient, however, to the extent that the employer seeks to draw upon the knowledge and expertise of the worker (Collins 2003, p. 229). In modern employment relationships, most employees have some discretion in the way they perform their tasks.

The employment relationship can therefore be modelled as an agency relationship. An agency relationship is a relationship between two people, one of whom (the ‘principal’) benefits when the other (the ‘agent’) performs some tasks. The employer-principal benefits more when the employee-agent takes care or uses effort when performing the task, than when the agent fails to take care or use effort. If the principal cannot directly observe the agent’s effort level, we have what is called an agency problem. The employee might engage in low effort, or what is often called ‘shirking’.

Employers want to minimize agency costs. Employees want this too, to the extent that they share in the gains. One way to reduce agency costs is to invest in monitoring. Employers do this all the time (Posner 2000): they buy computer programs that count the number of keystrokes, they videotape their cashiers, they listen to telephone operators, they hire supervisors. But monitoring is costly, so there is a limit on how much monitoring is profitable.

Another way to reduce agency costs is to design a contract that makes compensation depend on the output of the agent. Systems under which workers are paid for their output are piece-rate pay, payment by commissions, profit-sharing, and bonus plans. The problem with output-based pay systems is that they place employees at risk of having earnings that are variable over time. This is so because output does not merely depend on the worker’s input but also on external factors outside the control of the employee. When the employee is risk-averse – and this will almost always be the case – he will demand insurance against unlucky outcomes. The challenge is then to design a ‘mixed contract’ so that there are incentives to perform well, but without burdening the workers with too much risk.
Employment contracts

There are some further complications (Posner 2000). One complication is the multi-tasking problem. Professors, for example, have not only one task, but must perform a number of tasks: to give a certain number of lectures (task 1), with a certain quality (task 2), and to publish a certain number of articles (task 3), of a certain quality (task 4). If payment is made dependent only on the performance of one task, for example, on the number of articles published, the employee’s incentives to perform well with respect to the other tasks would be severely impeded. The solution is to measure the performance of all tasks, but the transaction costs of doing so may simply be too high. The employer may therefore decide to pay a flat fee in the hope that the employee will spend effort performing all tasks.

A particular problem with output-based pay exists when many employees have to cooperate in performing tasks. When output is really the result of teamwork, it is difficult to measure individual contributions to output. On the other hand, when tasks require cooperation among multiple agents, but agents are paid only on the basis of independent work, they will shirk on cooperation tasks. For example, piece-rate workers will work very fast on the assembly line but will not take the time to train newcomers.

A final complication is that often both parties to the employment contract will have agent-like and principal-like qualities. The employee might depend on the employer’s engaging in the right conduct. For example, an employee might be expected to achieve a certain target, but the employer is expected to provide the employee with some training. With respect to the target, the employee is the agent and the employer is the principal, but with respect to the training, the roles are reversed. If the employer fails to provide the training, the employee may find it too hard to perform. As a result, the optimal contract might split the profits between the two parties.

Given the difficulties with output-based pay plans, another method of increasing the chances that workers will not shirk their duties is to pay them a wage above the market wage. This method has been the object of the efficiency-wage theories. The general point of these theories is that ‘over-paying’ an agent is a way of reducing agent shirking or other misbehavior. Employees realize that even though supervision may not be detailed enough to detect shirking with certainty, if they are caught cheating on their promises to work hard and are fired as a result, the loss of a job paying above the market wage is costly. If an employee’s work is not diligent and he is fired, he faces the risk of earning a lower wage. Because workers want to keep the valuable job, they will work hard to avoid being dismissed. Employers can afford paying the higher wages (efficiency wages) because employees work harder than otherwise and firms can save on monitoring costs. Shapiro and Stiglitz (1984) state in their shirking model that if all
employers were to follow the strategy of raising wages, then the incentive not to shirk again disappears; the worst that can happen to a worker who shirks on the job is that he is fired, since he can be rehired (assuming there is no unemployment) at the same high wages. But as all firms raise their wages, supply of labor would exceed demand and unemployment would result. With unemployment, even if all firms pay the same high wages, a worker has an incentive not to shirk. For, if he is fired, an individual will not immediately obtain another job.

A related method of assuring faithful performance is the backloading of compensation (Posner 1995, p. 59). To increase the expected cost to the employee of being detected and fired, the employer may shift remuneration toward the end of the employee’s career, for example, in a generous pension. If an employee has an entitlement to a generous pension that he will lose if he is found guilty of malfeasance, he will have an incentive to behave himself. The pension is funded by the worker’s accepting a lower wage. In effect, he posts a bond, which he pays for by accepting the reduction in wage and which he forfeits if he misbehaves and is fired. Likewise, the employee could be underpaid (paid less than marginal product) early in his career and overpaid later on (Lazear 1979). Some of the employee’s compensation is then deferred as a reward for productive behavior. This results in an upward-sloping age-earnings profile throughout the worker’s career in the firm.

3 Termination rules and consequences of termination

A A taxonomy of dismissal doctrines

(i) At-will termination The at-will doctrine says that an employment contract of indefinite duration can be terminated by either party at any time, for no reason, good reason, or a bad reason. Employees have no legal or contractual ground to challenge the dismissal in court. What the rule basically says is that the employer does not have to give any public justification for his termination decision. The at-will doctrine confers an unfettered discretionary power to terminate the employment contract.

(ii) Good faith limitation on termination When courts apply the good faith restriction on termination, it is recognized that although the employer has the opportunity to terminate the employment contract unilaterally, such discretion does not include the privilege to engage in an opportunistic exercise of termination rights. In jurisdictions where a good faith standard is available, the courts still recognize a strong presumption in favor of an unfettered discretionary power to terminate. The good faith standard,
however, enables a court to control discretionary decisions that are perceived to be based on improper purposes, that is, where the power is used for a purpose not originally expected by the employee. In other jurisdictions, a similar function can be achieved by the doctrine of abuse of rights, under which a legal right is, in the view of the court, used for an improper purpose. When a good faith limitation is available, the employee will usually bear the burden to prove that the dismissal was not made in good faith. In some jurisdictions, the legislator has specified when dismissals are assumed not to be in good faith, for example by stating that dismissal is arbitrary when based on conduct wholly unconnected to the employee’s performance at work.\textsuperscript{1} The employee can then establish a prima-facie case for a lack of good faith which the employer must rebut by showing cause.

\(iii\) Just cause limitation on termination

Although different legal systems have their own particular formula for determining the justice and fairness of a dismissal, these abstract phrases such as ‘adequately justified’, ‘socially justified’, ‘reasonable’ or ‘just cause’ all imply that judges determine whether the dismissal is justified. Dismissals are presumed to be unreasonable unless the employer can prove that the dismissal was justified according to some measure of substantive and procedural fairness. Not only must the employer demonstrate a substantial reason for dismissal and the reasonableness of dismissal for that reason, but he or she must also persuade the courts that he or she conducted the process of making the dismissal fairly. Workers are permitted to contest the reasonableness of their dismissals, and if employers are not able to prove that they have a just cause for the dismissal, employees may be reinstated in their jobs or receive compensation.

The just-cause doctrine is in essence a choice of third-party governance: the employment contract can only be terminated if a third party (for example a judge) approves it. The courts are not merely reviewing the exercise of power but are substituting their own view of how power should have been exercised. It may happen that the employer considers an act of misconduct or lack of competence serious enough to warrant termination, but the court does not agree with this assessment based on the standard of reasonableness.

\(B\) The at-will versus just-cause debate

All of the above-mentioned rules can be observed in American as well as in European labor markets, but there is a tendency for stronger employment

\textsuperscript{1} See for example, the Belgian rule on arbitrary dismissal of blue collar workers (Article 63 of the Law on Employment Contracts).
protection regimes in Europe. The issue of debate in Europe is whether the rule for employment termination should be made less rigid in order to make European labor markets more flexible, whereas in America, the discussion is whether there is a need to give employees more protection in their job. The law and economic literature on the intrinsic merits and failures of different employment termination rules – the results of which will be summarized here – may make a valuable contribution to the ongoing debate.

(i) Administrative costs  Because employees have no legal or contractual ground to challenge dismissal in court under the at-will rule, neither party has to incur litigation costs upon termination of the contract. Under a just-cause rule, unfair discharge suits will frequently be initiated by the employee. Apart from the legal fees that need to be paid in litigation, the employer has to prove to the court that he has a just cause for terminating the employee. However, just cause does not simply mean that more evidence is necessary to support a claim that an employee is performing poorly. It means that poor performance must be proven to an outsider, so verifiable evidence will be needed (Morriss 1996, p. 1925). This implies that the employer will need to implement information-gathering and record-keeping procedures to document instances of misconduct in a verifiable way. But it also implies that an employer cannot act on observable but non-verifiable evidence. In many cases the most important element in the success of a business is cooperative effort, which depends heavily on attitude and morale. Yet these are often the most difficult elements to explain to an outsider (Epstein 1995, p. 160). The employer who knows his reasons for action might not be able to articulate them in ways that are persuasive to outsiders. The costs of a just-cause requirement include the information-gathering costs of documenting instances of misconduct and the direct and indirect (including the risk of error) costs of litigation. The risk of legal error exists because judges are asymmetrically informed about the employee’s performance and are insufficiently informed to determine whether or not the employee’s performance was adequate within the particular requirements of the job. When it is difficult to prove the inadequacy of the employee’s performance, it will be difficult for employers to terminate even inefficient employees without liability. When employees know this, their incentives to perform may well be diluted.

2 Note that employment parties in the US could always bargain for more employment protection in their individual employment contract and that employment parties in Europe can have employment contracts with very little job protection, such as when an employer contracts for work through a temporary help agency.
A somewhat related cost is the inherent uncertainty cost when judges give content to the standard of reasonableness (just cause) \textit{ex post} in the context of adjudication. Even when the employer successfully demonstrates a substantive reason for dismissal, courts will still consider whether or not it was reasonable for the employer to fire the employee for that reason. An employer might think at the time of firing that termination is the reasonable thing to do; however, courts might decide \textit{ex post} that it was not. This uncertainty regarding the consequence of his actions under a just-cause regime is costly for a risk-averse employer.

(ii) \textit{Implicit self-enforcing contracts} Another economic reason for the desirability of contracts at will comes from research involving repeated games. After the contract has continued for a while, one party may devise threats for the purpose of extracting the other’s gains. The theory of repeated games shows how the threat of dissolution can itself lead to a self-enforcing agreement. That the aggrieved party can easily walk away from the agreement after misbehavior by the other – as is the case under a contract at will – will sometimes serve as a sanction sufficient to deter opportunism. An implicit self-enforcing contract is one where opportunistic behavior is prevented by the threat of termination rather than by the threat of litigation (Klein 1980, p. 358). It is left to the judgment of the parties concerned to determine whether or not there has been a violation of the agreement. No third party intervenes to determine whether a violation has taken place or to estimate the damages that result from such violation. If one party violates the terms, then the only recourse of the other party is to terminate the agreement after he discovers the violation. Both parties continue to adhere to an agreement if and only if each gains more from adherence to, rather than from violation of, its terms (Telser 1980). Still, the circumstances under which the threat of walking alone would be a sufficient deterrent are quite severe: the value of the contract to both parties must either be expected to continue indefinitely or have a substantial positive probability of continuing in all future states (Rosen 1984). Otherwise, there are well-known tendencies toward opportunism as the end-period approaches, and the contract is no longer self-enforcing. This problem has been discussed in economics under the name of the ‘last-period’ problem. People keep promises because the prospective gains from doing so exceed the prospective losses.

But what if they have no prospects because of the proximity of retirement or switching jobs or some other mode of exit? One possible market response is the backloading of compensation. For example, if an employee has an entitlement to a generous pension that he will lose if he is found guilty of malfeasance, he will have an incentive to behave himself even if
he is in the last period of employment (Posner 1995, p. 59). But then again, the loss of potentially large amounts of deferred pay upon termination might appear to be an excessively severe, unfair sanction in comparison to the actual damages resulting from the employee’s deviation from expected performance.

A low detection rate accounts for the seeming paradox of a harsh penalty, dismissal, for relatively minor incidents of shirking. According to Rock and Wachter (1996, p. 1924), the explanation rests on an asymmetry of information: ‘workers know their work effort; employers do not. Firms can learn by monitoring, but constant monitoring is costly. To save on costs, firms infrequently monitor workers. If most shirking goes undetected because of high monitoring costs, firms must penalize employees an amount greater than the expected loss of any specific incident’. Benjamin Klein (1980, p. 359) states that – after recognizing the importance of transaction costs and the incomplete ‘relational’ nature of most real-world contracts – termination must be unfair in the sense that the actual penalty is greater than the gains from cheating: ‘Rather than the usually analyzed case of costlessly detected and policed contract breach, where the remedy of making the breaching party pay the costs of the damages of his specific breach makes economic sense, the sanction here must be large enough to make the expected net gain from shirking equal to zero’. However, it also has been recognized that the obvious concern with such seemingly unfair contractual arrangements is the possibility that the employer, under an at-will system, terminates the contract opportunistically without cause in order to get out of the obligation of deferred compensation. To this concern we turn next.

(iii) Opportunistic terminations The at-will rule vests an unrestricted discretionary power in employers to terminate the employment relationship. This power could therefore be used in an opportunistic way. It has been argued that, although the employer is not legally constrained in doing so, de facto constraints restrict his behavior. When workers and firms share the costs of specific investments, they both have incentives to maintain stable relationships. Recall that the exercise of the power to terminate affects the obligations of both parties, since once the power has been exercised, neither is bound to perform any further obligations under the contract. The employer who fires an employee has to pay an implicit price, that is, he will no longer be able to reap the benefits of the worker’s labour. Employers do not want to fire employees with specific investments for no reason, because to do so would inflict a loss not only on the employees but also on the firm itself. The employer immediately loses the investment in human capital that he has made in training the dismissed worker to become
Employment contracts

73

an effective part of the production process. In addition, he will have to incur the hiring and training costs associated with a replacement worker.

However, there are instances where the employer’s desire to obtain a return on his investments in the relationship with his worker will not inhibit him from discharging workers in an opportunistic way. Suppose the worker paid for the bulk of his firm-specific human capital by accepting a below-market wage in his early years, expecting to recoup it by an above-market wage in his later years. The employer can defeat the recoupment by firing him. After a point, a worker’s productivity usually declines with age, yet he may still be in his recoupment period – earning a wage that may exceed his current marginal product. Second, the employer may have back-loaded the employee’s wage as a method of assuring faithful performance. The employer has an incentive to fire the employee if by doing this he can get out of the obligation of deferred compensation.

Still, the employer might be restrained in doing so by reputational considerations. Employers have a desire to have a reputation as a ‘fair’ employer. If the employer gets a reputation for unfairly or arbitrarily discharging employees, he will have difficulty in attracting new workers or he will have to pay new employees a wage premium. However, reputation as a bonding equivalent is an uneasy and slippery notion. Where the risk of opportunistic termination is high, it may well be that a contract that allows the employer’s actions in terminating to be controlled by courts is efficient even when the additional enforcement costs are taken into account.

(iv) Conclusion Both the at-will rule and the just-cause rule have intrinsic merits and failures. The at-will rule saves on monitoring and enforcement costs but may add significant cost caused by opportunistic behavior. The just-cause rule protects the worker against opportunistic termination but is inherently costly to enforce. At-will contracts should therefore be observed when the enforcement costs of other non-self-enforcing contracts (like just-cause contracts) are sufficiently large in relation to the costs of opportunism associated with at-will contracts, whereas just-case contracts should be observed when the costs of opportunism in self-enforcing contracts (like the at-will contract) are large in relation to the administrative costs associated with just-cause contracts. The relative efficiency of contracts at-will versus just-cause contracts hence depends on the circumstances that dominate the labor-market exchange. These circumstances surrounding the employment exchange may vary across national labor markets (European labor market versus American labor market), and within national labor markets, across a variety of employment relationships (skilled labor versus unskilled labor and specific skills versus general skills) and they may even vary during the life-cycle of the employee at a particular firm. Schwab (1993, pp. 12–13)
suggests that the ‘variety of relationships may in itself counsel against a uniform legal approach to employment terminations’. Furthermore, he proposes a life-cycle just-cause approach that protects employees only in their more vulnerable stages of employment, namely at the beginning and the end of their careers.

C Obligations upon termination

(i) No damage compensation Just as a firm may need to attract new employees, it will sometimes need to reduce employment, either permanently because of long-term shifts in its demand for labor or temporarily because of transient shocks that affect productivity. Such separations are important for realizing labor-market efficiency. Similarly, the employee may find that the relationship is no longer advantageous. For example, the employee may find a better job elsewhere. When the employment parties draft their employment contract, they might know that the gains from exchange that induce the contract are of uncertain duration. In an employment contract, the worker may get a superior offer whose value exceeds that of the existing job. These offers tend to come in at random intervals. Similarly, the firm may suffer adverse fortunes that cannot be foreseen at the time the exchange is initiated. An exchange is efficient if and only if the gains from the exchange to both parties exceed their opportunity costs (Rosen 1984). The opportunity costs are the gains forsaken in order to complete the current exchange or, stated differently, the value of the next-best alternative to the current exchange.

From an efficiency point of view, the relationship should be ended when the total gains to both parties from continuing the relationship fall short of the total gains from termination. If the point in time at which the gains from the relationship vanish is foreseeable at the moment that parties draft their employment contract, then a definite duration can be written into the contract; but when the appropriate point for termination cannot be foreseen so precisely, the agreement must be open-ended or of an indefinite duration (Rosen 1984). If the employment parties were able to draft a fully contingent, complete contract, the contract of indefinite duration would specify termination whenever certain non-advantageous contingencies occur and specifies continuation otherwise. A solution of this kind would require that all possible contingencies are knowable in advance and easily verified ex post. In addition, specifying all possible contingencies, of which there may be millions, is very costly. For these reasons, it can be concluded that the fully contingent contract to achieve the efficient allocation of resources to their-highest valued use (that is, efficient continuation and efficient separation) is not possible due to excessively high transaction
costs. How then can an efficient allocation of resources be achieved when employment contracts of indefinite duration are necessarily incomplete?

In the contract law and economics literature, it has been explained that expectation damages provide parties with optimal incentives to perform when contracts are incomplete (Shavell 2004). If expectation damages apply when the contract is breached, parties will perform their contract when performance is efficient and will breach the contract and pay damages when breach is efficient. The question is whether expectation damages provide incentives to the parties in an employment contract of indefinite duration to continue the relationship when efficient and to terminate and pay damage compensation when termination would be efficient while at the same time saving on the transaction costs of writing a fully contingent complete contract. In other words, would employment parties benefit from a system under which the party who initiates the separation pays for the expectation loss of the other party?

According to Milgrom and Roberts (1992), there are several informational impediments to a solution of this kind. First, determining the appropriate size of the payment would be problematic because the idiosyncratic value of the relationship to each party is not likely to be freely known by the other or by the courts. Second, if different payments to be made depend on whether the separation is initiated by the employee (‘a quit’) or by the employer (‘a discharge’), then the distinction can quickly become blurred. In the words of Milgrom and Roberts (1992, p. 348):

An employer can often make an employee’s life so miserable at work that he or she just has to quit, or an employee can misbehave so badly that the employer sees no choice but to fire the offender, and yet third parties cannot tell who is to blame. This makes a separation payment system very problematic because it will not be clear who should pay whom.

Having different payments to be made dependent on who initiates the termination gives employment parties an incentive to play a ‘you quit first game’ (De Geest forthcoming). The party who contemplates terminating the relationship has an incentive to induce the other party to quit first in order to get out of his obligation to pay damages and to receive damages instead. Who has the best chance to win a ‘you quit first game’? The party who is best able to impose negative externalities upon the other party, for example by making the situation at work unbearable for the other one. The ‘innocent’ party will quit first, but has to pay damage compensation to the ‘guilty’ party under a system that makes the terminating party liable for damage compensation whereby courts are unable to verify whose actions really caused the dissolution of the relationship. In that case, it might be better to abandon the requirement to pay damage compensation upon
termination in order to avoid the mistake that damages are paid by the ‘innocent’ party.

De Geest et al. (2001) have given another reason for the absence of damages for employment contract termination. Under the expectation measure, the terminating party would have to compensate the other party for the loss of expected profits. In case of employee termination, the prospect of having to compensate the employer for his expected profits might cause the employee to reduce his effort level. Employers earn the highest profits on those workers whose productivity is high relative to their wages. If the employee’s past productivity relative to his wages is used as a proxy to determine the loss of expected profits for the employer when the employee leaves, then the result is that high-performing employees will have to pay higher damage compensation than low-performing employees when they leave the firm.

We know that employers draft contracts for employees to sign, but they usually expect employees to do much more than strictly required under the terms of the contract. The explicit contract only defines a verifiable but minimum standard of performance, like showing up for work during working hours. With monitoring being less than perfect, the employee has some freedom to decide how hard he will work. Incentive-based pay systems will to a certain extent motivate employees to work hard, but intrinsic motivation is at least as important for that purpose. For the efficiency of the labour exchange, it is important that these incentives are not destroyed. But expectation damages would have the opposite effect. If the employee knows that damages to be paid in the event of leaving increase with any increase in the level of supra-contractual performance, then he will choose not to do more than is strictly required under the explicit terms of the employment contract.

The conclusion is that a damage-based termination system is not a good idea for employment contracts. How then can the objective of efficient termination and efficient continuation be achieved when the use of expectation damages cannot be used for that purpose and employment contracts are incomplete and relational in nature? Some European legal systems have attempted to provide a judicial or administrative mechanism that requires the employer to justify to the public authorities the need to make economic dismissals. In The Netherlands, employers are prohibited by law from unilaterally terminating an employee without the express permission of the relevant administrative agency (Article 6 of the Extraordinary Decree on Labor Relations 1945). A termination permit is granted only after an official hearing by the administrative agency that has to determine whether appropriate grounds exist. The employer who seeks
as to allow termination without compensation when the total gains to both parties from continuing the relationship fall short of the total gains from termination and so as to refuse termination in the reverse case. However, judicial or administrative determination of whether an economic dismissal is justified is a problematic system for the reasons put forward by Collins (2003, p. 188):

> Numerous problems have been encountered by such mechanisms, not the least being the difficulty for public officials of making the requisite business judgement about the requirements of the business and the options available. How can a judge, for instance, assess the validity of the employer’s financial projections, assessment of the product market, and rejection of other alternatives? Even if a suitably qualified administrative body were established, it might not receive from the employer the detailed information needed to make a proper judgement, and employers might need to act more quickly than a complex administrative process permits. For these reasons, legal control of the substance of the decision that there is a need to make economic dismissals seems unworkable.

All of the above-mentioned problems increase the attractiveness of the termination at-will rule for the termination of an employment contract of indefinite duration. Under the at-will rule, none of the terminating parties has to pay damage compensation and no intervention by a third party is required. At the same time, parties economize on the costs of drafting and enforcing explicit and complete employment contracts. Does the at-will rule result in the allocation of resources to their most valuable use? It seems so. Under the at-will rule, the employer is entitled to terminate an indefinite duration contract when continuation on its present terms is no longer advantageous to him due to a change in circumstances. If termination in fact does not represent the joint-maximizing outcome considering the interests of both parties, the terminated employee would be free to negotiate a re-division of the joint contractual surplus in order to induce the employer not to terminate. The employee is equally entitled to terminate the contract when a better alternative job opportunity exists. However, the arrival of a better offer signals a need for negotiation and re-contracting so that the previous terms of the contract can be modified. About these renegotiations Rosen (1984, p. 984) states:

> These negotiations are always bilateral, because with full information and cooperation, a worker never quits unless the employer wants him to, and a worker

a termination permit must prove to the satisfaction of the agency that the reasons behind his intended termination are consistent with legal criteria that would make such a move reasonable.
is never fired unless it is beneficial to both parties for him to leave. A rational party never unilaterally walks away from an efficient contract. At-will contracts focus re-contracting to those points in time when it is beneficial to recontract, i.e., when the relevant states actually change, and may save resources for this reason.

However, the assumption that renegotiations are bilateral and only occur when an unforeseen event makes adjustments of the contract terms beneficial does not hold in all situations. Even when the relevant states have not actually changed, a party might opportunistically seek to modify the terms of the contract to his advantage when the other party has increased his vulnerability, for example by making heavy relationship-specific investments. This is why the parties might adopt notice periods, severance pay provisions, seniority rules, or covenants not to compete as methods of retaining some capacity to encourage specific investments within the flexible environment created by the contract at will.

(ii) Notice periods and severance pay

Most European courts and legislators apply to employment contracts of indefinite duration the implied term that only permits unilateral termination upon giving reasonable notice. Some legislators have specified mandatory minimum notice periods. The remedy for the employee for breach of the implied or mandatory notice period is confined to a claim for net wages during the notice period. Advance-notice provisions attempt to reduce or avoid periods of frictional unemployment. Frictional unemployment arises because labor markets are inherently dynamic, information flows in labor markets are imperfect, and because it takes time for unemployed workers and employers with job vacancies to find each other (Ehrenberg and Smith 1997, p. 568). Advance notice provides workers with the opportunity to search for new jobs prior to their displacement so that periods of unemployment after their displacement are reduced or avoided. The hope is that the employee will have found a suitable alternative job by the end of the notice period.

The implied notice period is not always the same for all workers. Such factors as seniority, age, wage, type of function or type of worker (blue-collar worker or white-collar worker) might be used to determine

4 WARN (Worker Adjustment and Retraining Notification Act 1988), for example, requires US employers of 100 or more workers to give employees and local government officials 60 days’ advance notice before they shut down or make large-scale layoffs.
the implied duration of the notice period.\textsuperscript{5} Is there an economic logic to applying different notice periods? If notice periods attempt to reduce or avoid periods of frictional unemployment that workers expect to suffer after displacement, then notice periods will vary according to the average time needed to find a suitable, alternative job.

The speed with which workers find (and accept) jobs depends, among other things, on the number of jobs offered in a given period in the relevant labor market. In thick labor markets, there are many job offers, whereas in thin labor markets, there are few job offers in any given period. For example, a lawyer who agrees to work for a law firm that specializes exclusively in negotiating oil-tanker mortgages might make it very difficult for himself, in the event he was ever discharged, to find an equally remunerative position (Posner 1992, p. 148). Although he could always find some kind of work in the market, the search to find a job at another firm that also specializes in oil-tanker mortgages is likely to be protracted because the job market for lawyers with that particular specialization is thin. But that is all the more reason why the lawyer might demand, as a condition of working for such a specialized firm, that it agrees that should it ever discharge him, it will continue to pay his salary during the notice period, the duration of which corresponds to the average time needed to find an equally remunerative position. In fact, by taking up a job for which a thin market exists, the lawyer, in the example, gives up alternative job opportunities, because he could have chosen to start working in a field of law for which more job opportunities exist. Foregone alternative job opportunities are a special form of reliance in employment contract. Notice periods protect this reliance to some extent (De Geest forthcoming).

De Geest (forthcoming) also explains how the distinction made in certain legal systems between white-collar and blue-collar workers with respect to the duration of the notice period they were entitled to, made sense in earlier times. There was a time in history when employment in the goods-producing and agricultural industries (with mainly blue-collar jobs) was much higher than employment in the service industry (with mainly

\textsuperscript{5} In Belgium, for example, a complex system for the calculation of notice periods exists. Different notice periods apply for blue-collar and white-collar workers. The latter category is in general entitled to longer notice periods. Within the category of white-collar workers, a distinction is made between lower-paid employees and higher-paid employees. When judges are required to determine the duration of higher-paid white-collar employees, they usually make use of a formula, the so-called formula Claey: notice period in months = \((0.87 \times \text{length of service}) + (0.06 \times \text{Age}) + (0.037 \times \text{remuneration} \times \text{index 2007/index month of dismissal}) - 1.45\).
white-collar jobs). Because it took white-collar workers on average more time to find a suitable job upon displacement than blue-collar workers, longer notice periods were needed for white-collar workers. But over time, the number of white-collar jobs has expanded whereas the number of blue-collar jobs has contracted (especially in the agricultural sector) due to changes in the industrial distribution of employment. Therefore the distinction between blue-collar and white-collar workers has become less relevant from an economic point of view, although some legal systems continue to make that distinction for the purpose of determining the duration of the notice period.

Meanwhile, changes in the production technology within each sector have also required that workers acquire new skills and work in new jobs. There has been a general increase in the demand for skilled workers, both in blue-collar jobs as well as in white-collar jobs. Investments in human capital (skills and knowledge) increase the worker’s productivity and hence the wages that the skilled employee obtains (after the training period). Some of the human capital is firm-specific, meaning that it increases the worker’s productivity, hence earnings, only in a particular employer’s employ. On average, it will take a longer time to find an equally remunerative job after displacement, because it is in general more difficult to find high-wage offers than low-wage offers. It is also possible that the employee will never be able to find a job that pays such high wages as he used to earn before. The factors of seniority and wage used for the calculation of the duration of notice periods account for the difficulties that specifically trained workers might encounter in finding suitable alternative jobs after displacement. Seniority can be used as a proxy for the amount of investment the employee has made in a particular job, because the employee is more likely to have gathered a greater amount of specific human capital as his length of stay at a particular firm increases. The wage level is used as a proxy for the amount of investment, because the higher the wage, the more likely it is that it contains an implicit return for firm-specific human capital.

Age may depreciate a person’s ability to find a new job. We know that employers are reluctant to hire older workers (Hutchens 1988). The cost of training an older worker is higher than that of training a younger one because of the age-related decline in fluid intelligence, while the expected return to the investment in training is lower because the older worker has a shorter working life expectancy (Posner 1995, p. 329). This ‘age effect’ is accounted for when longer notice periods are implied for older workers. In some jurisdictions, courts are reluctant to give too much weight to the age factor or do not take that factor into account at all. This is understandable, because the entitlement of older workers to longer notice periods adds to
the cost of employing older workers, and hence to the reluctance of future employers to employ them.

One reason why employers might be reluctant about notice periods is the so-called ‘last-period’ problem. The worker who knows that his employment will come to an end after the expiration of the notice period might lose his motivation to work hard, but worse, could cause large negative externalities, for example, by stealing or damaging company property, and criminal punishment for his malfeasance is not a realistic prospect. We know that employers often prefer to terminate the employment immediately, even if this implies that they have to pay damages in lieu of notice. An alternative solution is to adopt a severance pay provision. Some legal systems have a judicial or administrative system that entitles employees to severance pay when the employer terminates their contract unilaterally. The factors that play a role in the calculation of the amount of severance resemble those that are used for the determination of the duration of notice period, especially seniority and wage, and sometimes age.6

If one regards the threat of termination to be exclusively a device to motivate the performance of the employee, then notice periods or severance pay limit the penalty for non-performance and attenuate the impact of termination as a useful enforcement device. Legal systems deal with this concern to the extent that notice periods and severance pay are not unconditionally enforced. If the employee has committed a fundamental breach of contract, providing in that way an ‘urgent cause’ for termination, an employer is permitted to terminate the contract unilaterally and summarily without payment for any notice period whatsoever. Similarly, severance pay can be denied or reduced if the employee was seriously at fault in procuring the dissolution of the employment relationship.7

(iii) Covenants not to compete

Covenants not to compete forbid employees for a certain period of time after termination of their employment contract from entering into a contract with an employer operating in the same market or from starting their own enterprise in that market. The employer has an injunctive remedy available when the covenant is breached. This means that the employee can be enjoined – within the scope

6 In the Netherlands, judges of the subdistrict court award severance pay upon dissolution of employment according to the following formula: severance pay = A (length of service with higher weight to the service years of older workers) x B (monthly salary) x C (correction factor).

7 The C factor in the severance pay formula used in the Netherlands will be reduced when – in the opinion of the subdistrict judge – the employee is to blame for the dissolution of the employment relationship.
of the covenant not to compete – from working for another employer. In most legal systems, covenants not to compete are treated with suspicion. Legal and judicial restrictions apply to their enforcement.

One economic rationale for using covenants not to compete in employment contracts is to protect the employer’s investments in the production of confidential information, such as trade secrets and customer goodwill, which is costly to produce but can be disseminated at low cost. The other economic rationale for covenants not to compete is to protect an employer’s investments in an employee’s general human capital. Without the use of a covenant not to compete, the employer would have no protection against the employee using the general human capital for which the employer paid to obtain a higher wage from another employer. Of course, parties could write a specific duration into their contract – during which the employee cannot quit without having to pay damage compensation – as a way of reassuring the employer that he will earn a return on his investment, but covenants not to compete protect employers against the appropriation of their investments even in a flexible environment like the contract at will. Unlike the contract of definite duration, which directly enforces the performance of the contract, the covenant not to compete is an indirect form of contract enforcement. The employee can quit without liability and will only be sanctioned when he starts working for another employer in violation of the terms of the covenant not to compete.

An economic reason, put forward by Posner and Triantis (2001), for why it might be efficient to regulate and restrict the use of covenants not to compete is that covenants not to compete might enable employers to externalize upon third parties (part of) the costs of investment in specific human capital. When the employer does not internalize all the costs of such investments, his ex ante incentives to make specific investments might be excessive. The possibility of externalizing costs exists when the former employment parties – usually together with a future employer – negotiate an agreement that releases the employee from the covenant in return for a payment. Such an agreement is important in realizing labor market efficiency when the total gains from employment in a new job are higher than those in the former job. In the release agreement, the former employer will at least require a payment sufficient to compensate for the loss of surplus that he expected to receive from his former employee. To the extent that specific investments increase the amount of the payment, the former employer will be able to externalize the costs of his investment to the new employer. When the present employer realizes that – with the adoption of a covenant not to compete – he will be able to externalize the costs of his specific investments in bargaining with the future employer of his employee, he will over-invest in specific human capital. This inefficient
use of resources is a basis on efficiency grounds for the court’s interference with the employment parties’ freedom to include a covenant not to compete in their employment agreement.

4 Pre-contractual information problems

A Pre-contractual information duties

Asymmetric information problems exist in the market for employment contracts just as in other markets. Employees have superior information on a number of characteristics that affect their productivity on the job, such as schooling, past performance, motivation and their health. Employers have better information on the content of the job, career opportunities and prospects of the firm. Asymmetric information hinders efficient job matching and should be corrected. At the pre-contractual stage, when parties are negotiating, they have an opportunity to exchange information. However, in real-world employment contract negotiations, both parties might decide to remain silent or even lie about some aspects. The question is whether such behavior entitles the disadvantaged party to a right to claim a legal remedy, such as the dissolution of the employment contract or damage compensation on the basis of the doctrines of mistake, fraud or related doctrines. Whereas from a legal point of view, the presence or absence of such elements as bad faith, dishonesty, lying or being silent matter for the outcome of litigation, from an economic point of view, rules are analyzed in the light of the incentives they give to produce and reveal information.

The central insight from the law and economics literature on pre-contractual information is that efficient rules have the effect that information is produced and revealed by the least-cost information gatherer (Kronman 1978). Applied to employment contract negotiations, this means that rules should be applied in such a way as to give the employee incentives to produce and reveal information on his personal characteristics because the employee is the least-cost information gatherer in this matter. Likewise, the employer is the least-cost information gatherer for matters concerning the job and the business. In legal terms, this means that from an efficiency point of view, employment parties should have a claim for compensation or dissolution of the employment contract when one of the parties remained silent or was dishonest with respect to information affecting the potential value of the exchange and for which he or she is the cheapest information gatherer.

However, an efficiency case can be made for a number of exceptions to the least-cost information gatherer rule (De Geest et al. 1999). First, information should only be produced and/or revealed when the information is sufficiently relevant to the other party. Requiring the least-cost information

Gerrit De Geest - 10.4337/9781781953068.00011
Downloaded from Elgar Online at 11/26/2018 09:08:09AM via free access
gatherer to spend resources on producing and revealing irrelevant information is a waste. An efficiency case can even be made for giving the employment parties a ‘right to lie’ on matters that are irrelevant in assessing the value of the exchange. Relieving employment parties of a positive duty to give information on irrelevant matters but not from the duty not to lie on irrelevant matters when asked would allow the other party to use the dishonesty as a pretext for claiming dissolution or compensation when he or she comes to regret the agreement for reasons that have nothing to do with the dishonesty of the other party. Of course, it needs to be determined which information is considered to be relevant. For the employee, information is relevant when it pertains to remuneration in a broad sense, such as wages, working situation, certainty, etc. For the employer, information is relevant when it is related to the productivity of the employee.

Second, parties should not have a duty to reveal entrepreneurial information. It might happen that one of the employment parties possesses information that was costly to produce, but easy to distribute among other interested parties. In order to avoid free-ridership and maintain incentives to create productive information, parties should have the right to remain silent and to lie on entrepreneurial information. Third, parties should be free to make statements that are non-verifiable either because they are subjective in nature or because no generally accepted definition of the statement exists.

The following rules are implied by an efficiency analysis of the kind made above. Employees have the right to lie on those health aspects which do not reduce their labor productivity in the short and medium run. However, they should honestly reveal information on health conditions which affect their productivity in the short or medium run, like pregnancy. Employees do not have to reveal information on their criminal record unless it makes them unfit to do a particular job. Employees should be honest about verifiable facts that predict their productivity, such as diplomas and the wage they earned in their former job. However, employees are free to make non-verifiable statements about their capabilities, even if they know that they may not be not true. For example, the employee may say ‘I am a hard worker’, or ‘My English is excellent’ or ‘I am the best employee’. There are no generally accepted criteria to assess the validity of these statements; they are subjective in nature. There is also no way to assess whether the employee, in making these statements, truly believed them. Hence, these non-verifiable statements should not become the subject of litigation. The employer should be honest about verifiable facts like wages and other benefits for the first period of employment but should be free to make unverifiable statements such as ‘There is a very pleasant atmosphere in the firm.’
When it comes to existing laws, legal systems sometimes require no duty to honestly reveal information where an efficiency analysis would come to the opposite conclusion. This is for example the case with laws protecting pregnant women by giving them a right to lie about their pregnancy.

B Probation periods
In legal systems with employment protection legislation, it is common for employment parties to insert into the employment contract a probation period clause. Employment protection legislation does not apply during the probation period, but usually the agreed upon probationary period may not extend beyond a legally imposed maximum duration.

De Geest (forthcoming) argues that the probation period is a relatively costly technique of acquiring information about the capacities of the employees, on the one hand, and on remuneration in a broad sense, on the other hand. To the extent that the information is already available to one of the parties before the start of the employment contract, the first-best solution is to induce the informed party to reveal that information along the lines developed in the previous paragraph. Probation periods are also costly in another way. When employment protection is considered to be necessary to induce parties to make relationship-specific investments, the fact that it is temporarily not enforceable during the probation period might make parties inefficiently delay such investments.

C Discriminatory signals
Employers know little about job applicants. In choosing among them, employers rely on signals to predict performance (Spence, 1973). For example, a job applicant with a university degree can easily provide the employer with a transcript. The university degree signals traits like intelligence that the employer values, but which he cannot directly observe at a job interview. The original models of job-market signalling concerned the ‘rat-race’ that could arise when the signal has no intrinsic value. Individuals might then over-invest in signalling.

Another type of signal is the discriminatory signal. In discriminatory signalling, a fixed trait, like gender, race or age, is used as a signal for an unobserved variable (Cooter 1994). The problem with discriminatory signals is not that they result in over-investment but rather that they may reduce the incentives of the ‘victims’ of the discriminatory signal to invest in their human capital. Discriminatory signalling is also known as statistical discrimination: the reliance by decision-makers on various statistical proxies to judge unobserved characteristics. Where non-discrimination laws apply, they prohibit the use of discriminatory signals like age, race and sex.

However, there is a transaction cost explanation for the use of
discriminatory signals. The use of a single, readily determinable characteristic as the basis for a hiring decision economizes on the cost of information. To illustrate, men are physically stronger than women on average, so some employers reject all female applicants for jobs requiring strength. By adopting such policies, an employer will often make mistakes like rejecting a strong woman and accepting a weak man. If these mistakes cost less than gathering more individualized information, the use of the signal maximizes profits, and competition will reinforce the discriminatory practice (Cooter 1994).

The usual objection against the use of discriminatory signals is that it is unfair towards the atypical group members who do not track the average experience of their group. Due to statistical discrimination, some employees will be treated much worse than their true characteristics justify. This in turn will distort the incentives of the above-average members of the group to optimally invest in human capital, since they will be simply treated as an average member of their class. This is true, but a legal ban on the use of discriminatory signals, such as age, gender or race, is usually not the cheapest way to undo their effects, because a ban forces employers to rely upon a more costly substitute for every job applicant. When the costs of making an individualized assessment of an employee’s performance are prohibitive, as they often are, prohibiting the use of particular proxies will lead to the substitution of other proxies. According to Posner (1995, p. 327), the problem of under-investment that may exist when age is used as a proxy in employment will then be shifted, not solved: ‘Whoever is “unfairly” disadvantaged by the new proxy, in the sense that it does not measure his abilities accurately (…) will lack the incentive to make the optimal investment in his human capital. And if the new proxies are less efficient – as they probably will be, because otherwise they would in all likelihood have been adopted without government prodding – wages will fall because employer’s labor cost will be higher, and with lower wages there will be less incentive for workers to invest in their human capital’.

A cheaper solution than banning the use of discriminatory signals, according to Cooter (1994), consists of augmenting the information flow about potential victims of discrimination in employment. Those individual employees with characteristics that justify better conditions could provide that information to the employer, provided that such information could be produced and credibly communicated. The objection to this form of market remedy against statistical discrimination is that those individuals who make use of it will need to incur additional costs to inform the employer. This objection could be overcome, according to Cooter (1994), where the state provides additional information on employees without charge.
5 Other possible failures of individual bargaining

According to the economic theory of market regulation applied to employment contract regulation, an economic case for regulating the content of the employment contract requires demonstration of market failure and a policy to correct it. It is assumed that parties will bargain for efficient employment contracts in perfect markets. Efficient terms are those of which the costs to the person who promises to perform are lower than the value of that performance to the other party. For judicial or regulatory intervention to be appropriate from an economic point of view, a specific market failure must be identified. A large part of the law and economics literature seeks to provide economic grounds for contract regulation. Such grounds include unequal bargaining power like monopoly power, externalities, public goods and information problems.

Are there reasons to assume that there will be inefficient employment contracting?

For labor lawyers, the basic argument for legal regulation of the content of the employment contract is the notion of inequality of bargaining power. It is true that workers sometimes will enter into bargains that do not work out well ex post, but it is extremely unlikely that workers will be systematically exploited in any meaningful sense (Fischel 1984). Inequality of bargaining power in the sense that the employer is a monopsonist would be a market failure, except that employers are rarely monopsonists and even if they were, it could not account for the prevalence of inefficient non-wage terms in employment contracts\(^8\) (Freed and Polsby, 1989). Work is a necessity for most workers, but workers are a necessity for employers too, because a non-filled job is costly. Moreover, the availability of unemployment benefits and higher family incomes reduces the need for employees to accept the first job that is offered to them. It is common practice among employers to hand the employee a standard form contract that sets out the respective obligations of the parties. However, standard contract terms – those offered on a take-it-or-leave-it basis with no opportunity for bargaining – are not sufficiently

---

\(^8\) Rational monopsonists act in ways that maximize their profits. Whereas the monopsonistic employer will reduce wages and employment below the competitive level, the non-wage terms of the employment contract will tend to be efficient. Employment conditions will be included in the contract with the monopsonistic employer, where the costs of providing them are lower than the value of the condition to the employee. The employer will act in this way because it allows him to offer even lower wages to the employee. The wage decrease compensates the employer for the cost of providing the benefit, and employees accept the decrease because the value of the benefit outweighs its costs to him.
suspect to satisfy the unequal bargaining proposition. The standard form contract is a practice that reduces transaction costs by avoiding the costs of negotiating and drafting separate agreements with each employee. Standard form contracts will not contain onerous terms when employers compete for workers. Competing employers offer employees various packages of benefits at different prices. From this angle of differing opportunities, bargaining may be said to have occurred implicitly. It is not necessary that all workers in a particular market be well-informed on all opportunities for the market to behave competitively. So long as there is a sufficient number of informed shoppers, firms will not be able to offer inferior conditions of employment or pay below-market wages (Schwartz and Wilde 1979).

It is however possible to isolate particular types of employment risks for which it is unlikely that a sufficient number of employees will be perfectly informed. The typical examples are occupational safety and health risks. A perfectly informed employee would have perfect knowledge about the expected accident or illness costs. Under the process of compensating wage differentials, employers who offer more dangerous jobs would have to pay higher wages than those who offer safer jobs in order to compensate the employee for the increased risk. Therefore, employers have optimal incentives to take safety precautions because this will enable them to attract workers at lower wages. This analysis is changed when information problems are considered. Workers may have no good knowledge about workplace dangers because these are often so subtle that the costs of informing the workers are prohibitive. This implies that compensating wage differentials are inefficiently small and workplace safety too low. This is an economic argument for making the employer liable for workplace accidents, as is the case in many legal systems.

Another economic rationale for regulation of employment contract terms is the existence of harmful externalities. The adoption of particular contract terms might impose losses upon third parties and transaction costs may preclude them from internalizing these costs. For example, society as a whole would suffer losses if the employer had an unrestricted right to terminate the employment contract of an employee who refused to perform unlawful acts or reported an illegal activity (whistleblowing). Schwab (1996) explains that much wrongful discharge law can be understood as a search by the courts for these third-party effects.

6 Conclusion
The recognition that most employment contracts are relational in nature is important for our understanding of the specific rules of employment law, for then the general law of contract is unsuitable for handling all disputes
that may arise in connection with employment contracts. Relational contracts involve relationship-specific investments made by one or both employment parties, the impossibility for the parties to draft a fully contingent complete contract due to uncertainty about the future, the difficulty of proving contractual breach in a verifiable way to a third-party enforcer such as a court, and the difficulty for the employment parties of observing contractual breach due to asymmetric information. Given these typical features of the employment contract, the economic role of employment law is to provide a legal framework for the maximization of the joint value of the labor exchange by reducing the employment parties’ incentive to take advantage of unspecified or unenforceable elements of the contractual relationship while economizing on the costs of writing and enforcing detailed employment contracts. This could be done in two ways: a legal framework to foster self-enforcing implicit contracts or a legal framework for third-party governance. Under third-party governance, courts or arbitrators will decide on important adjustments of the employment contract, including its termination, ex post when more information about the states of the world is available. Under a self-enforcing implicit contract, parties secure cooperation by the threat of termination at will when the other violates the implicit terms of the contract; third parties do not intervene to determine whether a violation of the terms of the implicit agreement has taken place. The optimal choice between these two systems depends on such circumstances as the level of information available to third parties to make appropriate decisions and the extent to which the danger of opportunism may make the contract at will no longer self-enforcing. For reasons of completeness, it should be noted that some legal systems apply an intermediate form of employment contract governance. The starting point of the intermediate legal system is that employment contracts are at will, vesting in employers discretionary powers to adjust and terminate the contract; the good faith standard, however, enables a court to control discretionary decisions that are perceived to be based on improper purposes. It is also common for intermediate legal systems to read into employment contracts such terms as notice periods or severance pay, as methods of retaining some capacity to encourage specific investments within the flexible environment created by the contract at will.

Another economic role of employment law is to correct failures in the market for employment contracts. The employer’s liability for workplace accidents, for example, has been explained as serving this goal. Finally, some suggestions have been made for ways in which legislators can intervene to improve the efficient functioning of labor markets. Legislators can increase the flow of information on workers who would otherwise suffer from the use of discriminatory signals. Legislators can also reduce
the potential for opportunism in employment contracts by increasing the alternative job opportunities available to workers; this can be done by standardizing production methods and by making labor markets larger through the removal of mobility barriers.

**Bibliography**


Knight, Frank (1921), Risk, Uncertainty and Profit, Boston: Houghton Mifflin.

Gerrit De Geest - 10.4337/9781781953068.00011
Downloaded from Elgar Online at 11/26/2018 09:08:09AM
via free access
Labor and employment law and economics


3 Regulating unions and collective bargaining
Kenneth G. Dau-Schmidt and Arthur R. Traynor

1 Introduction
Historically, developed nations have undertaken the regulation of unions and collective bargaining in the pursuit of three over-arching objectives: to foster unions and collective bargaining; to minimize industrial strife; and to promote union democracy. Nations have pursued these objectives not merely because these ends themselves were viewed as laudable, but also because it was thought that they served underlying societal goals of equity, efficiency and voice (Budd 2004).

Proponents of the right to organize have argued that collective bargaining promotes equity in bargaining power between labor and management and allows employees to gain a larger share of the fruits of their efforts. Such a redistribution of wealth is desirable not only because it produces a more equitable distribution of the rewards of production, but also because it promotes economic growth and political stability. Workers deserve a greater share of the proceeds of their labor and this income makes them better workers and consumers and adds strength to the fiber of our democracy (Atleson 1983, pp. 35–43).

However, in regulating labor relations countries generally go beyond merely protecting the right to organize. Labor laws prescribe and proscribe various behaviors in the selection of collective representatives, the conduct of collective bargaining and the enforcement of collective agreements. The proponents of such regulation reason that, in order to obtain the greatest net benefit from collective bargaining, it should be conducted under a system that promotes the rational selection of collective representatives, useful exchanges of information between labor and management, ‘good faith’ bargaining and the efficient enforcement of collective agreements. Such regulation is thought to improve the outcomes of collective negotiations and minimize the costs of negotiation and enforcement. Economists have argued that employee voice can improve economic efficiency in the negotiation and enforcement of employment contracts, the reduction of turnover costs and the management of companies (Freeman and Medoff 1984). Moreover, although the occasional resort to economic warfare is necessary for the private resolution of disputes through collective