Donado, Alejandro; Wälde, Klaus

Working Paper

How trade unions increase welfare

CESifo working paper: Labour Markets, No. 3618

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Donado, Alejandro; Wälde, Klaus (2011) : How trade unions increase welfare, CESifo working paper: Labour Markets, No. 3618, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:
http://hdl.handle.net/10419/52469

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
How Trade Unions Increase Welfare

Alejandro Donado
Klaus Wälde

CESifo Working Paper No. 3618
Category 4: Labour Markets
October 2011

An electronic version of the paper may be downloaded
- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp
How Trade Unions Increase Welfare

Abstract

Historically, worker movements have played a crucial role in making workplaces safer. Firms traditionally oppose better health standards. According to our interpretation, workplace safety is costly for firms but increases the average health of workers and thereby the aggregate labour supply. A laissez-faire approach in which firms set safety standards is suboptimal as workers are not fully informed of health risks associated with jobs. Safety standards set by better-informed trade unions are output and welfare increasing.

JEL-Code: J510, J810.
Keywords: occupational health and safety, trade unions, welfare.

Alejandro Donado
University of Würzburg
Department of Economics
97070 Würzburg
Germany
alejandro.donado@uni-wuerzburg.de

Klaus Wälde
Mainz School of Management and Economics
University of Mainz
Jakob-Welder-Weg 4
Germany – 55131 Mainz
klaus@waelde.com

October 2011
We would like to thank numerous seminar (including Louvain-la-Neuve, Heidelberg and Yale) and conference (including conferences of the Royal Economic Society, EEA, EALE and CESifo) participants for lively discussions and comments, especially Raouf Boucekkine, Bill Brainard, Guido Cozzi, Matthias Doepke, Switgard Feuerstein, Michael Kaganovich, Jürgen Meckl, Giuseppe Moscarini, Paul Segerstrom, Hylke Vandenbussche, Fabrizio Zilibotti and an anonymous referee. We are especially indebted to the editor for very constructive comments. This paper and the companion paper Donado and Wälde (2010a) are extended versions of an earlier version circulated as Donado and Wälde (2008).
1 Introduction

The process of economic development and growth is a process of an endless introduction of new technologies. This is especially true for the early stages of the Industrial Revolution but also applies today. When new technologies are introduced, their properties are not always well understood. While a technology might promise that a certain good is provided very efficiently, the same technology could also have side effects that did not occur to the inventor. The history of the introduction of new technologies is full of countless examples.

Since as early as the Roman Empire, coal has been used as a source of energy. Systematic coal mining, however, was not carried out until the Industrial Revolution, when a massive and steady supply of energy was required. Coal seemed like the perfect solution. Mining, however, has its side effects. In 1831, a potential causal link between working in a coal mine and black lung disease was first reported by a Scottish physician. Nowadays, black lung disease is accepted as a disease caused by the repeated and year-long inhalation of small amounts of coal dust. However, it took more than 130 years for this link to be generally accepted. Only in the 1960s, after extensive political activities by various worker groups in Pennsylvania, Ohio, and West Virginia on the Appalachian coal fields, was black lung disease recognized as an occupational disease. As a consequence, the Coal Mine Health and Safety Act was passed in 1969 which established more comprehensive rules for work conditions and also the compensation of disabled mine workers (Smith, 1987).

There is an abundance of further examples of worker movements improving health and safety conditions, including “brown lung” disease caused by exposure to cotton dust (Botsch, 1993), “white lung” disease caused, inter alia, by mining and the exposure to asbestos (Rosner and Markovitz, 1991), the health risk posed by radium (Clark, 1997), workplace exposure to dibromochloropropane, a pesticide that makes workers sterile and is linked to the risk of cancer (Robinson, 1991), the spray machine conflict in the early 1900s (Frounfelker, 2006) or conflicts in the pottery industry (Stern, 2003) and in the automobile and steel industries (Bacow, 1980, ch. 5). For an overview of the literature on the history of occupational health and safety (OHS), see Judkins (1986, p. 240). A more general history of labor standards with international comparisons is covered by Engerman (2003).

A reading of these analyses shows that the side effects caused by new ways of production reveal themselves only gradually. While there might be uncertainty about health implications of a certain job, there is initially often simply ignorance about health implications, sometimes just absence of any doubt. When workers then start sensing that “something is going wrong”, that work conditions are causing health problems, these claims are often met with doubt, not only by employers, but also by insurance companies or even the government. These analyses also clearly demonstrate that worker movements, joint collective actions by individuals, are required to raise political awareness, to lobby for changes in work conditions and to eventually bring about regulatory changes towards better OHS measures.

Similar conclusions about the importance of worker movements for triggering broader
support not only for the improvement of working conditions but also for the development of the modern welfare state can be drawn when looking at Germany. During the Industrial Revolution around 1850, the issues of poverty, working and living conditions of dependent workers caused organizations to be created enabling workers to express their own interests (see e.g. Schneider, 2005, p. 15). While poverty and dependent work also existed in pre-industrial times, the contemporaneous rise of the wealthiness of some and the poverty of others was no longer accepted as “the will of God”. The first trade union in Germany, founded in June 1848 by type setters, was set up with an aim to secure the living standards of type setters, who feared competition from the steam engine and technological progress (hence, there was income orientation), but also with an aim to establish mutual health and disability insurance systems (Schneider, 2005, p. 27). The worker movement, represented by unions and political parties, was also incited by occupational injuries which almost caused “mass causalities” (Tennstedt et al. 1993, p. XXI), partly due to the widespread use of new technologies and fast economic growth. These movements and associated political pressure caused Bismarck, the German chancellor, to implement, inter alia, statutory accident insurance in 1884.

The outcome of this discussion about historical episodes in advanced OECD countries is threefold: (i) A safe workplace, in short OHS, does not come for free: Achievements of the modern welfare state, which today are taken for granted, were hotly disputed in the past. (ii) There is a conflict of interest between unions and firms, which goes beyond pure wage bill issues. In many cases, industry, insurance companies and often also the government initially object to any demands for compensation or changes in health standards simply because there is no clear scientific medical evidence for the claimed nexus between certain symptoms and the professional activity. (iii) Unions\(^2\) played a crucial role in pushing for OHS standards and prepared and fought for what is (almost generally) accepted today as a positive aspect of modern welfare states (see e.g. Brugiavini et al. (2001, ch. II.2.1), Agell (1999, p. F144) and the discussion below). Only once workers succeed in forming large groups and in lobbying for their joint interests is there enough political visibility in order for changes in OHS regulations to take place. To put it briefly, in the spirit of Freeman and Medoff’s (1984) “collective voice”: Trade unions have a “good face” as well.

The purpose of this paper is to understand why it took worker movements (rather than the government or employers) to start the development of insurance mechanisms. Why did worker movements eventually lead to the creation of government agencies which regulate OHS nowadays and what are the determinants of endogenous OHS standards?

We shall construct a model which highlights the key ingredients for understanding the importance of worker movements in the past. Jobs have two effects on workers - they provide income and they affect health. In order to keep the analysis as simple as possible, we will assume that workers are entirely ignorant about the health implications of jobs: job choice

\(^2\)We will often use ‘union’ as name for more informal worker groups, worker movements or worker associations. Union, as used here, does not necessarily describe a well-organized and at times bureaucratic institution as is nowadays the case in some OECD countries.
is based purely on the wage paid by the employer. Returning to the coal miner example from above, workers were simply not aware of the potential risk of black lung disease. Given the absence of any information on the health risk of working, the production process exerts a negative externality on workers' health. This translates into a pure health externality affecting utility of workers and into an aggregate labour supply externality. OHS standards can in principle reduce these externalities but they also reduce the total factor productivity (TFP) of firms, reflecting the fact that OHS is costly.

We embed this view into an economy with one homogenous good and inelastic aggregate labour supply (before taking the health externality into account). We assume perfect competition on goods and labour markets, implying, inter alia, full employment. Unions do not cause unemployment in our setup. Firms hire workers from a spot market. As countries in early stages of development do hardly exhibit any specific worker-firm attachments (like long-term contracts or health insurance systems), firms can simply replace a sick worker by a healthy one. As long as health effects of working are disputed, no employer or government would concede better working conditions. The role of worker movements is to provide and confirm information about the health effects of working. An individual worker does not have enough time and makes too few observations to discern job-related health effects from other health effects. A group of workers, a union, has many members and thereby more observations. Learning is much faster and unions can thereby help internalize the externality.

In standard trade union models, the objective of trade unions is to maximize the wage income of their members. We extend this arguably narrow perspective and portray trade unions as having both high wage income and good health standards as their objective. We then find determinants of OHS standards by letting unions set OHS standards. This monopoly view of OHS-setting unions and employment-setting firms is - as in wage-setting models of unions - a short-cut to a more complete setup with endogenous union membership where workers form groups to increase the speed of learning.

Some of our findings are as follows: Each firm individually is opposed to higher OHS standards as they reduce TFP and thereby profits. Unlike compensating differentials setups with complete information, competitive markets here are unable to take health effects caused by technologies into account: individuals can not judge with sufficiently high precision to what extent a certain job will affect the health. The laissez-faire factor allocation is characterized

\[ \text{We see this complete ignorance as a short-cut to a Bayesian learning setup where workers form a prior about health implications and it takes time to learn the true health consequences of a job. See Viscusi (1979, 1980) for various applications of Bayesian learning to uncertainty about health implications of jobs.} \]

\[ \text{In her account of the transformation of the US labour market in the twentieth century, Goldin (2000) argues that labour markets never where 'spot markets' as portrayed in theory. But she also writes that “such labor markets had attributes far more characteristic of “spot” markets than do labor markets today” (p. 550). We therefore believe that the assumption of a spot market (in the theory sense) is the best modeling choice for real world markets at that time.} \]
by inefficiently high sickness leaves. The externalities at work are a health externality and an aggregate labour-supply externality. The former affects both healthy and thereby employed workers and workers currently without job.

If better-informed firm-level trade unions set OHS standards, the positive effect on the improved health of their members balances the negative effect of lower employment due to lower TFP. If there are economy-wide or occupational unions, OHS standards are more comprehensive as unions also take the negative health effect on overall labour supply into account. Whether a union confederation succeeds in internalizing all externalities depends on membership rules to be discussed further below. As both firm-level unions and union confederations internalize at least some of the health and labour-supply externality, the presence of unions is output- and welfare-increasing.

The paper starts with a short review of the related literature. Section 3 presents the model. Section 4 analyses the difference between the centralized and decentralized equilibrium. The reason for suboptimal decentralized equilibrium is presented and it is shown why majority voting would not resolve the issue. Section 5 argues that firm-level trade unions move safety measure towards their socially optimal levels and that union federations set close to welfare-maximizing safety levels. Section 6 applies this framework to understand the dynamics of trade union density from 1880 to 2008 for selected OECD countries. An emphasis is put on cross-country differences in preferences - inspired by American Exceptionalism. The final section concludes.

## 2 Related literature

Our paper is related to various strands of literature. First, there is obviously a huge literature on trade unions, and it would be impossible to provide a summary here which does any justice to the various substrands. While it seems fair to argue that most contributions attribute a distorting (efficiency-reducing) role to unions, there are also some economists that find positive aspects in union behaviour: Brugiavini et al. (2001, ch. II.2.1) see unions as the precursor to the modern welfare state. They write on p. 163 that “unions developed mutual insurance as part of associational self-help to compensate for the lack of private insurance or public social protection. At the same time, they mobilized [...] for the expansion of social rights. Increasingly, many of the protective functions that unions provided [...] came to be taken over by the state”. A by now well-accepted argument was made by Freeman and Medoff

---

5 Distortions can have their positive sides in second-best worlds or when it comes to collecting rents. See Mezzetti and Dinopoulos (1991) for an example with an employment-oriented union in an international trade setup with imperfect competition.

6 Historical evidence linking union growth to their provision of insurance (strikes, unemployment, sickness, burial cost) for the Netherlands and Britain is provided by van Leeuwen (1997). Quantitative evidence for the United States for union decline due to an expanding welfare state is provided by Neumann and Rissman (1984).
(1984): By providing a “collective voice”, unions provide information which otherwise would not be available. Malcomson (1983) argues that unions increase efficiency as they improve the allocation of risk-bearing between firms and workers. Acemoglu et al. (2001) argue that unions induce training and provide insurance. Booth and Chatterji (1998) and Viscusi (1979, ch. 11) show how trade union bargaining with monopsonistic firms increases social welfare and Agell (1999, p. F144), more generally, argues that “certain institutions may serve quite useful purposes” in the labour market. We put forward OHS standards as an example of such a useful institution. We believe that this beneficial historical aspect of worker movements for what are now modern societies and the role unions can play in developing countries today has not received sufficient credit so far. Our contribution lies in the emphasis and analysis, in the framework of a very simple model, of the informational and learning advantage of a union in a world with incomplete information and side effects caused by new technologies.

Second, and maybe most importantly, our view of multi-feature workplaces is related to but differs starkly from the equalizing differences approach of Rosen (1974, 1986). Equalizing differences are traditionally derived in setups with perfect information. When workers know about all job characteristics and all markets are competitive, factor allocation is efficient and any institution would be distorting. Given the historical situation and technological examples we have in mind, workers having perfect information does not appear to be a realistic assumption. We therefore choose the other extreme and assume that workers are unable to learn anything about work-related health implications. While the reality certainly lies somewhere in between, the justification for our assumption is simple: When new technologies become available, workers and often society as a whole do not know a lot about potential side effects. Health implications may only become apparent over the long-term and workers might simply not have the time to learn about these implications. Hence, with regard to learning processes which take a very long time, we assume right away that it is impossible for the individual workers to learn of health effects. As a consequence, a decentralized factor allocation is inefficient. In contrast, trade unions consisting of a large number of workers have access to many observations about jobs, can collect this information and can therefore learn more easily. In fact, we assume that unions have perfect information and can therefore internalize externalities, increase efficiency, output and welfare.

A companion paper (Donado and Wälde, 2010a) qualitatively and quantitatively studies the effect of globalization for labour standards in the North and in the South in the presence of unions as portrayed here.

3 The model

Our economy produces a homogenous good. Aggregate output amounts to $Y$. A typical firm produces the quantity $y$ by employing a certain number of workers $l$. All firms use the same
technology with total factor productivity (TFP) $A(s)$,

$$y = A(s)f(l).$$

The function $f(l)$ is characterized by positive but decreasing marginal productivity, $f' > 0$ and $f'' < 0$. This amounts to saying that $y$ has constant returns to scale as there is some other factor of production (capital, human capital of an entrepreneur or other) which is normalized to one.\(^7\) Given our historical perspective on what are now OECD countries or our focus on developing countries today, we assume that firms can hire from a spot market. There are no hiring or firing costs and it does not take any time to find a worker.

The central focus of this paper is OHS. This aspect is reflected in the production process in the TFP component $A(s)$. TFP in a firm or in a country is influenced by many factors, starting from very technology-specific aspects (like the management and communication skills of staff) and ranging to more economy-wide influences (like the institutional stability, the political regime, or the education level of workers). The more important factor influencing TFP for our arguments is OHS $s$. A job is safe(r) if a worker is (more) certain to return home in good health after 8 (or more) hours of work. We reflect safer jobs by a higher $s > 0$.

Safe workplaces are clearly in the interest of the worker, and in many cases, OHS is also a central concern for employers. If safety measures increase the smoothness of a production process, employers should be in favour of high safety standards. An accident in a coal mine, costing not only the lives of workers but also letting the production process break down for weeks, is clearly not in the interest of the firm. In many cases, however, there is a fundamental conflict of interest. In the case of low-skill workers or workers needing only general (i.e. not firm-specific) human capital to perform their job and in countries where firms do not (have to) pay sickness-leave (i.e. whenever firms can easily replace their workers), firms have no economic interest in the state of health of their workers. Quite to the contrary, OHS measures are costly. A workplace where coal miners are well protected against black lung disease or ore miners against silicosis is more costly than one without protection measures like ventilation systems. A worker who spends half an hour dressing and undressing (helmets, safety glasses, gloves, entire suits etc.) is less productive than a worker who starts doing his job right away.

What matters for our results is that workers value safety more than firms. For modelling purposes, we go to the extreme and exclude firms from any benefits from higher safety. We capture safety costs by letting OHS measures reduce TFP, $A'(s) < 0$.\(^8\) Given the spot market assumption, a sick worker would simply be replaced by a new healthy worker.

---

\(^7\)This follows the tradition in this literature (see e.g. McDonald and Solow, 1981, Oswald, 1982 or Booth, 1995). For an analysis with explicit constant returns to scale and capital as the second factor of production, see the discussion paper version of this paper (Donado and Wälde, 2010b).

\(^8\)This is the standard assumption in the literature on compensating differentials, see e.g. Rosen (1986). If $A$ increased in $s$, no uncertain jobs would ever be observed. One can always imagine that $A$ initially increases in $s$ but decreases above some threshold level. It could be that low $s$ reduces labour productivity rather than TFP. For simplicity, we will continue to use the term TFP.
An individual values consumption $c$ and health $z$ and both are determined by the job an individual chooses. A job is therefore a differentiated good as in Rosen (1974). Let $z(s)$ denote the share of days that an individual is healthy and can work. Currie and Madrian (1999) summarize the literature on health and labour markets. They document a positive relationship between health and income with health having a larger effect on hours than on wages. While it is true that the link between health and labour market participation is less clear-cut (Currie and Madrian stress that this could be due to an abundance of methodological problems), in the following we feel safe to assume that safety measures $s$ improve health, $z(s) > 0$. Utility of workers,

$$u = u(c, z(s)),$$

increases in consumption $c$ and health $z(s)$ per se, i.e. individuals value the share of days they are healthy, but with a decreasing slope, $u_c > 0$, $u_{cc} < 0$ and $u_z > 0$, $u_{zz} < 0$ (subscripts denote partial derivatives). Health is important for two reasons: It matters per se and consumption rises due to more working days per year. All workers are identical in their preferences.

On the aggregate level, consumption equals output $C = Y$ and labour demand $L$ equals labour supply,

$$L = z(s) N.$$

The latter is given by potential employment $N$ (a headcount assumed to be fixed) multiplied by the share $z(s)$ of, say, days a worker is healthy and can actually work. In equilibrium, this amounts to $L$ workers being healthy and working and $N - L$ workers being sick. Improved safety, implying improved health, implies higher labour supply.

We finally turn to trade unions. Depending on the degree of centralization of negotiations and wage setting, the literature usually classifies countries in three groups (see e.g. Calmfors and Driffill, 1988): (1) highly decentralized systems with wage setting at the firm level (i.e. USA and Canada), (2) intermediate degree of centralization (most continental European countries), and (3) highly centralized systems with wage setting at the national level (i.e. Nordic countries and Austria). We will also consider different degrees of centralization and model the two polar cases of highly decentralized and highly centralized systems.

In a decentralized setup, unions operate at the firm level and are therefore small in comparison to the economy as a whole. As we view spot markets as the best description of labour markets for activities as described in the introduction, there is no attachment of workers to the firm. Hence, membership of firm-level unions is just as volatile as employment at the firm. As a consequence, the union only cares about the overall well-being of the $l$ workers in this particular firm. As households value consumption and health, we let unions value these quantities as well. In fact, we make the union utilitarian and obtain, given (2), $lu(c, z(s))$. 

8
Given that employed workers earn full income, their consumption equals the wage,\(^9\)

\[ v = lu(w, z(s)). \tag{4} \]

In some countries, unions are large or form a confederation. Their basic objectives are the same but they now represent not only the workers of a particular firm but all workers that are currently employed,

\[ V = Lu(w, z(s)). \tag{5} \]

The main difference compared to the firm-level union is that health now has two positive channels, as in individual preferences (2): health matters per se and through higher labour supply visible here through \(L\). An alternative to economy-wide unions, also captured by (5), are occupation-specific unions. As long as a union takes the effect of standards on all workers into account (e.g. because a union represents all coal miners and not just those currently employed in one particular firm), beneficial labour supply effects as a result of higher standards are internalized by the union.

Both the small union and the confederation have a narrow membership concept. As only those that work in a firm are members of the union, workers that are sick are not taken into account. This is the difference between unions and a central planner which will play a role when discussing the welfare effects of unions.

4 Centralized and decentralized OHS setting

This section compares the behaviour of a planner to the setting of OHS levels in a decentralized economy. This allows us to understand the basic mechanism of why trade unions in principle can have positive welfare and output effects.

4.1 The decentralized economy

The standard view to a setup with multiple job characteristics is Rosen’s (1974, 1986) equalizing-differences approach. According to this approach, workers enjoy (or dislike) job characteristics in addition to the wage and a worker’s utility function would look like the one we use in (2). The difference to our approach consists in the criteria for choosing a job. In the equalizing-differences approach, workers have full information about job characteristics and the choice of jobs would depend both on health implications \(z(s)\) and on income leading to a consumption level \(c\). Firms can therefore choose wage-safety pairs on a worker’s indifference curve. The resulting market equilibrium would be efficient.

\(^9\)For an introduction to the discussion on the appropriate specification of union preferences, see Oswald (1982) and Booth (1995, ch. 4). Note that even for modern Britain, there is evidence that physical working conditions is one important issue over which trade unions and management bargain (Millward et al., 1992, pp. 249-254).
The crucial difference from our approach lies in our historical perspective of unions in contemporary OECD countries and the conclusions we draw about information. Workers do not have sufficient information (neither would society as a whole) to perfectly evaluate the impact of work, a certain job or a specific technology on health. Workers could form expectations but their expectations need to be - in the absence of perfect information - based on a prior in a Bayesian learning sense. Perfectly competitive firms taking a safety-wage trade-off into account would then set an inefficient safety level if the prior is not identical to the true distribution of the health impact of a job. When on the job, workers would of course gradually learn about health implications of work, but each single worker makes just a few observations, especially when health also depends on other factors than just work and certain health impacts come with a long delay or can not easily be observed (as the examples in the introduction have shown). There is simply not enough variation; econometrically speaking, there is not a sufficiently large number of observations to draw firm conclusions and learning can take more than a lifetime. To capture this idea in the simplest way possible, we assume here that workers choose employment based only on the wage and firms choose employment taking the wage rate as given. This will qualitatively imply the same type of inefficiency one would observe in a Bayesian setup (as employed e.g. by Viscusi, 1979, 1980). The advantage of this shortcut is clearly the much simpler analytical tractability.

Given this focus of workers on wages, optimal firm behaviour yields the familiar equality between marginal productivities and factor rewards,

$$w = A(s) f'(l).$$

In a laissez-faire economy, a firm fixes the stock of labour and the safety level $s$. The derivative of profits with respect to the safety level is $d\pi/ds = A'(s)$, i.e. it is negative. Firms only see the TFP-reducing impact of increased safety. As a consequence, firms would like OHS standards to be as low as possible. The comparison point to the central planner solution $s^U$ or $s^Y$, to be characterised shortly, is a laissez-faire safety level of $s^*$. Given that we exclude negative safety levels, we can set $s^*$ to zero (or to the level where $A(s)$ starts to fall, see fn. 8).

4.2 The planner

The planner can take two guises. Either the planner maximizes aggregate output, or welfare. As all firms use the same technologies, aggregate output is obtained by replacing labour input

---

10 An alternative way to describe this is to let workers maximize utility (2) considering $z(s)$ as exogenous and identical for all. We would like to thank a referee for pointing this out.

11 Fishback (1998) argues that some compensating differentials seem to have been paid in the US around the beginning of the 20th century, but the “differentials rarely fully covered the expected losses” (p. 723). Our assumption of an absence of compensating differentials therefore appears to be a sufficiently good description of labour markets in earlier stages of development.
in individual output (1) by the labour-market equilibrium condition (3),

\[ Y(s) = A(s) f(z(s)N). \]  

(7)

A utilitarian welfare function takes into account that a worker can be in two states. When healthy and employed, utility is given by \( u(w(s), z(s)) \), where consumption in (2) was replaced by wage income. Denoting the latter by \( w(s) \), we stress the dependence of the wage on safety levels \( s \). When the worker is unable to work and labour market income equals zero, utility is given by \( u(0, z(s)). \)

Social welfare is therefore defined by

\[ U(s) = Lu(w(s), z(s)) + (N - L) u(0, z(s)). \]  

(8)

We see immediately from this structure that the safety level \( s^U \) chosen by the planner will generally differ from the one by a large union (5) as the latter ignores workers that are not employed. Whether this difference is of a technical nature or real depends on the determinants of union membership. Whether workers remain members of a union once they become temporarily or permanently unable to work depends on various institutional details that differ across countries and over time. We will return to this issue when we discuss our results.

If the planner focuses only on output maximization, the optimality condition giving the output-maximizing safety level \( s^Y \) can directly be seen from the aggregate production function (7),

\[ A'(s) f(z(s)N) + Af'(z(s)N) z'(s)N = 0. \]  

(9)

This condition balances the output-decreasing effect due to lower TFP, the first term, with the output-increasing effect due to more labour supply, the second term. Whether there is in fact an interior solution or not will depend on the functional forms of \( z(s) \) and \( A(s) \). Introducing some type of Inada conditions for \( z(s) \) and \( A(s) \) at \( s = 0 \) and \( s = \infty \), respectively, and letting the functions in (9) be sufficiently smooth would make sure that an interior solution actually exists. Conditions in this vein would imply that the left-hand side of (9) is positive for \( s = 0 \) and negative for \( s = \infty \).

A social planner maximizing welfare (8) will take the effect of the OHS level \( s \) on labour supply as in (3) into account. The wage rate is the competitive one from (6). Given that \( z(s) \)

\(^{12}\)It is well-known that this is not a completely innocuous aggregation. As labour is characterized by decreasing marginal productivity, output of the economy increases in the number of firms. We circumvent this problem by normalizing the measure of firms to one. As is also easy to understand, none of our results depends on this normalization.

\(^{13}\)One might want to argue that the inability to work implies that health should be smaller than \( z(s) \). This could easily be achieved by multiplying \( z(s) \) by a factor smaller than one. As this would not affect any of our results, however, we omit this factor.

\(^{14}\)Providing examples for functional forms (e.g. of the Cobb-Douglas or CES type) that yield an interior solution is straightforward (Donado and Walde, 2010b).
denotes the share of days that an individual is healthy and can work and her average health level, the per-capita welfare function reads

\[ U(s)/N = z(s)u(w(s), z(s)) + [1 - z(s)] u(0, z(s)). \]  

(10)

Simplifying notation by \( u(w(s), z(s)) \equiv u \) and \( u(0, z(s)) \equiv \bar{u}, \) the socially optimal safety level \( s^{U} \) is described by

\[ z'(s)[u - \bar{u}] + z(s)[u_w w'(s) + u_z z'(s)] + [1 - z(s)] \bar{u}_z z'(s) = 0. \]  

(11)

Condition (11) balances the welfare-increasing and welfare-decreasing effects of increased safety. The first term displays an increase in welfare as the share of healthy workers increases and as utility of healthy workers exceeds utility of sick workers, \( u > \bar{u}. \) The second term shows how employed workers lose as their wage falls and how they gain as they value health per se. The third term displays a gain again, reflecting the fact that sick workers also value more health. The only downside of higher OHS, the reduction of wages, stems from two sources. Recall that aggregate labour supply, as employed also in the aggregate production function (7), is given by \( z(s) N. \) Hence, the wage can from (6) be written as \( w(s) = A(s) f'(z(s) N). \) The first source of a wage drop due to higher safety is the drop in TFP. Second, more safety increases the labour supply at the aggregate level (since more workers are healthy), which in turn reduces marginal productivity.

Compared to the utility maximizing condition (11), the output maximizer misses two channels: While the negative effect through falling TFP and the positive effect through higher labour supply are taken into account, the direct effects on health, both of the employed and unemployed workers, is missing. One would therefore expect that the welfare maximizing health level is higher than the output maximizing level, \( s^{U} > s^{Y}. \)

Compared to the decentralized equilibrium, we see that there are two externalities at work here. First, the health externality stemming from the fact that workers value health per se. This externality affects both healthy workers being currently employed and sick workers. Second, there is a labour supply externality caused by an increase of aggregate labour supply if workers become more healthy. The health externality operates within firms and can potentially be internalized by firm-level unions. The labour supply externality operating across firms can only be internalised at the country level. We will see now to which extent a government is able to internalise these two externalities and further below what role unions can play.

4.3 The role of the government

Occupational health and safety standards in OECD countries are nowadays set by government agencies. Viewing these agencies as the outcome of some political aggregation of preferences, could one not argue that unions are not needed and a voting mechanism would be sufficient for
achieving socially desirable OHS levels? We will briefly inquire into this issue by presenting a standard median voter model. We will see that voting by uninformed workers yields a suboptimal OHS level.

A policy maker that seeks to maximize the number of votes commits to a safety level that corresponds to the safety level preferred by the median voter. The median voter can be seen either as an expected utility maximizer or as an individual that spends a certain share of his or her time as working and being healthy and the remaining share as being sick. In a full-information equilibrium, utility of individual $i$ would be given by

$$u_i = z(s) u(w(s), z(s)) + (1 - z(s)) u(0, z(s)).$$

In this type of equilibrium, voters are aware of the fact that OHS levels affect both the wage level and the share of time he or she can work. Formally, the voter is aware that $z = z(s)$. As the structure of (12) is identical to the per-capita social welfare function (10), this yields a preferred safety level identical to the one chosen by the planner in (11). As all individuals are the same, voting under full information would imply that the socially desired safety level $s^U$ is implemented. The one central shortcoming of this approach, just as with the compensating differentials view, is the assumption of full information.

Under the empirically more convincing assumption of imperfect information and the necessity for learning, individuals are not aware of the safety to health link. Individuals perceive health to be a function of nature, nurture or individual habits but not necessarily as the true relationship $z(s)$. Hence, an individual voter, not sharing information with other voters who have experience about potential work-health links, would maximize

$$u_i = zu(w(s), z) + (1 - z) u(0, z).$$

Individual health $z$ is perceived to be unrelated to OHS levels and only the effect of OHS levels on the wage is taken into account. A voting equilibrium would imply that a suboptimally low safety level would be set by the policy maker. In fact, one would expect a safety level $s^\pi$ as in the suboptimal decentralized equilibrium.

Summarizing, a government setting OHS levels via a voting scheme can do so successfully only if voters are fully informed. As historical examples show, this full information is not a convincing assumption. Voting would therefore aggregate individual choices, but these choices are still biased as long as information is not available. An exchange of information is needed and we will now see how this can be achieved through unions.

5 Endogenous OHS with trade unions

The previous section showed that both a laissez-faire equilibrium and a voting mechanism do not internalize any of the externalities. We will now show that trade unions reduce or
even eliminate the distorting effect. Why does the union have the knowledge and means required to do so? There are two reasons: First, unions have many members and the more members there are, the easier it is to learn about job characteristics beyond the wage. Due to its size, the union can collect information more easily than individuals. Second, in contrast to a loose group of workers that have no institutional connection, unions have the means to “prove” the link between bad work conditions and health. They can monitor the credibility of individual claims about work conditions more easily\(^{15}\) and they also have the power to impose better working conditions. Unions are a means to overcome the information and credibility problem of individual workers (see, for example, Fenn and Ashby (2004, p. 46) and Robinson (1991, pp. 41-7).\(^{16}\) In what follows, we will analyse the principles of optimal union behaviour both for a decentralized system (firm-level unions) and a centralized system (trade union confederation).

### 5.1 Firm-level unions

In the traditional monopoly union model (see Dunlop, 1944, Oswald, 1982), unions set the wage, firms choose employment and unemployment is the inefficient equilibrium outcome. We give unions market power as well, assuming that it is beneficial for workers to join a union, that unions succeed in learning about the work-health link better than workers and that unions succeed in solving the monitoring problem.\(^{17}\) This is our highly condensed version of historical processes: Historically, worker movements do not have any market power when they start. Political parties are often the vehicle through which public attention and support increase. If new regulations then improve OHS standards, they are put into force by the government. Indirectly, however, these new regulations are set by worker movements and this is what we capture here. Unions use their market power not to set wages - as in the traditional model - but to set the safety level \(s\). While unions in the real world are concerned with several issues of which wage negotiation is an important one, we focus here entirely on union activities related to improving work conditions as described in the introduction. Wages are perfectly flexible in our setup and there is no unemployment.

At the firm level, employment \(l\) in the union’s objective function (4) is given by the firm’s

\(^{15}\)The importance of unions in alleviating moral hazard problems has already been stressed by Beveridge in 1909 (quote taken from van Leeuwen, 1997, p. 786). Beveridge claims that unions of his time were in the best position to monitor the appropriate use of unemployment benefit payments.

\(^{16}\)Firms can also learn faster than individual workers as a firm hires many workers. Once the firm has learned about negative health effects of a certain technology, however, it might not be in the firm’s interest to reveal this information as workers with health problems that were incurred in the past could then file claims.

\(^{17}\)Giving unions market power allows us to use the elegant monopoly union setup. This should not suggest, however, that we make a second-best world argument where one distortion (the market power of unions) corrects for another distortion (imperfect knowledge). Unions are beneficial even without (or despite) market power as they provide a superior (collective) learning technology in comparison to individualistic learning. Future work could use a Bayesian learning setup where collective information collection alone improves welfare.
labour demand from (6) which, through TFP, is a function of the safety level, \( l = l(A(s)) \). The aggregate wage rate \( w \) in the labour demand function \( l(.) \) is taken as parametric by the firm-level union. The choice of the safety level \( s^u \) is perceived by the union to affect labour demand through TFP and health \( z(s) \). We can therefore describe the union’s objective by \( v(s) = l(A(s)) u(w, z(s)) \). Assuming an interior solution, the union’s optimal OHS level \( s^u \) is given by
\[
I_A A'(s) u(w, z(s)) + I_u z'(s) = 0. \tag{13}
\]

As in the planner’s trade-off, safety has a positive as well as a negative effect. The negative effect captured by the first term comes from the reduction of labour demand by the firm as a result of the cost associated with a higher level of safety. The positive effect, shown by the second term, is the direct effect of improved health on utility of workers in this firm.

The differences between the union’s optimal \( s^u \) from (13) and the planner’s \( s^U \) from (11) stem from three sources: The firm-level union does not take the effect on workers not working in this particular firm into account. This is why the term \( z'(s) [u - \bar{u}] \) in (11) is missing here. It does also neglect workers which are currently unable to work, i.e. \( [1 - z] \bar{u} z'(s) \) is missing. Finally, the detrimental effect of OHS levels is evaluated with respect to labour demand by this firm and not with respect to the equilibrium wage level. The firm-level union does, however, internalize the health externality for healthy workers. In this sense, it improves upon the decentralized equilibrium.

The firm-level union could internalize the entire health externality, i.e. also for sick workers, if workers remained members of the union also after having been replaced. Whether this will be so depends essentially on membership details of unions. If membership extends beyond the period where a worker is hired by a firm, a union could take well-being of former employed workers currently sick into account as well. If union membership ends at the moment of the end of occupation (as portrayed above), the union would not represent the interests of workers off the job. In the latter case, firm-level unions would not be able to internalize all health externalities. As a firm-level union is small, it could in any case never internalize the aggregate labour supply externality.

### 5.2 The trade union confederation

The union confederation has the same objectives as the firm-level union only that it represents not only the workers from a particular firm but the whole labour force. Consequently, employment in the union confederation’s objective function (5) is economy-wide labour supply \( L = z(s)N \) and the wage rate from (6) is the general equilibrium wage level, \( w = w(A(s), z(s)N) \). The safety level set by the confederation is denoted by \( s^V \). Writing the per-capita objective function as \( V/N = z(s) u(w(s), z(s)) \), the optimality condition is
\[
z'(s) u + z(s) [u_w w'(s) + u_z z'(s)] = 0. \tag{14}
\]
The optimality condition (14) features two positive effects of higher safety, the increase in the share of healthy workers, \( z'(s) \), and the increase in utility due to higher safety and thereby higher health. Losses result from the lower wage. As discussed after the optimality condition of the welfare planner in (11), these losses are due to a reduction in TFP and an increase in aggregate labour supply.

The two terms missing in (14) relative to the welfare planner condition (11) refer to workers currently sick. As the union confederation cares about workers only, the effect of higher safety levels is overestimated as the marginal health effect \( z'(s) \) is multiplied by \( u \) in (14), in contrast to \([u - \bar{u}] z'(s)\) in (11). By contrast, as the union confederation does not take the beneficial health effects for those being currently without job into account, the \([1 - z]\bar{u}z'(s)\) term in (11), the effect of higher safety is underestimated. Overall, it is therefore not clear without specific functional forms and parameter values, which effect will dominate.

It is obvious from this discussion that a union confederation does internalize the health externality for healthy workers and the labour supply externality. Just as the firm-level union, it does not internalize the health externality for sick workers as long as the latter do not remain union members. If a union confederation took the interest of sick workers into account as well, no market intervention of a planner is needed as the union confederation would internalize all externalities.\(^{18}\)

6 The rise and fall of trade unions across nations

Measures of union membership (either headcounts or densities) seem to suggest that there is an inverse U-shape observable for basically all OECD countries. After an initial rise at the end of the 19th century and at the beginning of the 20th century, measures of trade union size tend to decrease. While there is heterogeneity across countries, this seems to be the overall pattern.

This section first presents an impression of this overall pattern. It then uses our framework to provide an interpretation to the general time-series pattern. Finally, by adding some aspects related to 'American Exceptionalism', it tries to account for cross-country differences in this general pattern.

6.1 The big picture

Figure 1 provides an impression of union densities for a selection of OECD countries. Density is defined as union membership divided by employment. The historical data for Australia, Canada, Denmark, Germany, Norway, Sweden and United Kingdom come from Bain and Price (1980), the historical US data is from Freeman (1998), data for Austria, France, Italy

\(^{18}\)Normalizing utility when sick to zero, \( u(0, z(s)) = 0 \), would make the union confederation social welfare maximizing. But this would be more of a technical solution assuming away the issue of membership.
and the Netherlands come from Visser (1989). Data as of 1960 is from OECD (2011). While the historical data needs to be considered with care (see Visser, 1989, the appendix of Freeman, 1994 and Bain and Price, 1980 for discussions of data quality), the inverted U-shape is clearly visible.

The figure displays predictions of union densities based on a regression of densities on time, time-squared and a post 1959 dummy to capture the switch to OECD data. All coefficients are highly significant and $R^2$s are high. The figure reveals three features of union density across countries: First, the inverted U-shape seems universal. Second, the peak differs across countries in its level and, third, in its year.  

![Graphs by country](image)

**Figure 1** The inverted-U shape of trade union density from 1880 to 2008 in selected OECD countries

19 Whenever there is both data from the OECD and data from older sources, the more recent OECD source was used. Density in Visser (1989) is provided on a net and gross basis. We chose the series which provided more historical information. The basic time-series properties are not affected by this admittedly opportunistic choice.

20 These are not some artefacts of our estimation. This is also visible from the raw data. The data plus the STATA codes are available at www.waelde.com/UnionDensity.
6.2 The rise and fall and American Exceptionalism

How can our framework help to understand these patterns? Let us first focus on the inverted-U shape by, say, considering only one specific country. The Industrial Revolution amounts to a sequence of introductions of new technologies. These technologies initially have unknown effects on health. Individuals learn about health effects over time and form groups in order to share information. With an expanding industrial sector and an introduction of new technologies, a lot of learning about side-effects of technologies is required. As we argue above, collective action is important and unions emerge.

To capture union density, one can easily imagine a generalisation of our model where union members and non-union members would coexist. This would happen if (i) only union members experience higher safety levels and (ii) there is heterogeneity in individual costs in joining a union. These costs can be thought of as psychic costs arising from differences in attitudes towards joining groups. With higher safety levels yielding higher utility, only those individuals would join a union - ceteris paribus - whose preference of individualism is below some threshold level. Given such a coexistence of union and non-union members, a series of new technologies being introduced as a consequence (or characterizing) the Industrial Revolution implies a rise of unions over time.

If learning is successful over time and safety standards are being adopted by the government (see e.g. Neumann and Rissman, 1984, for empirical support for the US), the importance of unions declines. If the technologies that have been introduced in recent decades are less risky (given that health investigations are much tighter today than in the past), a continuing introduction of technologies does not require a continuing increase of union activity. Successful collection of information convincing the public of the necessity of better standards is the beginning of the end of unions. Success abolishes the very reason for the existence of safety-setting unions. The necessity of collective action in the first place and the adoption of union activities by government agencies as a consequence explains the inverted U-shape.

---

21 Such a model would be in accordance with existing theories (see Checchi and Lucifora, 2002, for a short but highly informative overview) where union membership is plagued by a free-rider problem if coverage by union wages is 100%. This problem is solved either by positive reputation effects for, or provision of private goods to, union members. Here, there is no free-rider problem as the positive effect of joining the union is higher individual safety and the negative effect is the cost of joining.

22 This does not discard other explanations for union growth as put forth e.g. by Freeman (1998), Checchi and Lucifora (2002) and many others. Both our benefits and our costs would be complementary to more standard determinants like institutional factors, business cycle states or composition of the labour force of the economy.

23 Another reason for declining importance of unions could be the fact that firm attachment of workers becomes more long-term - as it did according to Goldin (2002). When workers have firm-specific skills and can not easily be replaced, firms have a stronger incentive to provide better safety standards. Taking both views into account would require a model that allows a transition from spot markets to more stable worker-firm relationships. For a short overview on deunionization, see e.g. Acemoglu et al (2001) or Checchi and Lucifora (2002).
Let us now turn to cross-country differences in the U-shape such as level and year of the peak. Individual psychic costs for joining unions are crucial for understanding these differences and this is where American Exceptionalism (AE) comes into play: The origin of AE can be attributed to a statement in 1630 by John Winthrop, governor of the Massachusetts Bay Colony, stressing the exceptional circumstances American settlers find themselves in (Ferrie, 2005). An overview of the discussion about AE is provided by Voss (1993). According to her, AE manifests in the fact that workers in the US “were unable to build either a lasting broad-based labor movement or a powerful socialist party” (p.2). In this vein, Friedman (2010) relates AE to the fact that the US at some point was “the only country with no labor movement” and “the first country with declining union density”. In other words, the “American soul” is opposed to public institutions (Lucore, 1989) and even to collective arrangements as this restricts individuals “in shaping their own lives” (Ferrie, 2005, p. 199).

We now condense American Exceptionalism for a union-nonunion membership model into one single parameter, high psychic costs in joining a union. Costs are high in the US, compared to other countries, as unions resemble public institutions and restrict individual freedom. In other words, the US is a more individualistic society than many other societies. It is then a straightforward implication that the US had such a low and early peak. Generalizing this view to other countries would then imply that - ceteris paribus - the ranking of these countries concerning the date and the level of peaks of union membership densities can be linked to the distribution of costs of joining a group. Going from union membership to government regulation, more individualistic countries should also be characterized by less severe and less widespread regulations than countries with less individualism.

Summarizing, assuming that the health effects of the Industrial Revolution are the same around the world, union density would grow faster and to a higher peak in a country where the average preference for individualism is lower. American Exceptionalism explains the slow rise and slow peak of union density in the US. It should also imply that the general amount of regulations is smaller than in other countries.

---

24 Freeman (1994) inquires into the existence of “universal” effects of unions, i.e. effects a union has in any country. Freeman argues that AE can be found in the unions’ effect on wages as the union-nonunion wage gaps is the largest for the US. Another good example for AE is the fact that union coverage almost everywhere exceeds union density, but hardly so in the US (OECD, 2004, chart 3.4).

25 Various data sources offer information on attitudes or personal characteristics. Examples include the Hofstede index (Hofstede 2001, 2010) of individualism vs collectivism, various indices from personality psychology like the “Big 5” (Schmitt et al., 2007) and the World Values Survey. The latter was fruitfully exploited by Alesina, Glaeser and Sacerdote (2001) in their analysis of the differences in attitudes between the US and Europe. See also Schnabel and Wagner (2007) for an analysis of the impact of attitudinal variables on union membership.
7 Conclusion

The starting point of this paper was the belief that institutions like trade unions, which have been around for more than a century and are active in almost all countries in the world, are not just detrimental to economic production and welfare of a society. Studying activities of workers' associations and trade unions beyond wage negotiation has shown that trade unions play a major role in providing workplace safety - at least in providing information about the necessity of measures that assure occupational health and safety (OHS). Trade unions did perform this role historically in what are now OECD countries and do play such a role today in certain industrializing economies.

Can these OHS activities of unions assign unions an output and welfare increasing role? The crucial shortcoming of decentralized economies with health externalities from hazardous jobs is the lack of information about health effects. This lack results from the fact that an individual is generally not able to distinguish between various causes of good or bad health. An individual worker therefore does not have sufficiently many independent observations during his work-life to infer the effect of a job on health. This poses the question who is in a position to collect the information needed to understand health effects of working in a systematic way.

Inspired by many historical examples, where in early stages of development of an industry worker movements raised awareness about negative health implications of various new technologies, this paper argues that it takes a group of workers - called a trade union for simplicity - to collect this information. Worker movements that share a common goal exchange information, i.e. they "collect data" that allows then and them to identify health implications of a job.

In a world without exchange of information, a government would be unable to perform this job. A government seen as an executive body for an aggregate opinion obtained e.g. through majority voting would fail in implementing the appropriate safety measures. Individual voters in the absence of exchange of information would be misinformed and vote for a suboptimal level of safety measures. In this sense, this paper provides a rationale for unions based on their ability to collect information and derive conclusions about which type of jobs or technologies have negative health implications and which do not. A large union is able to internalize (at least partly) the direct health externality on workers and the aggregate labour supply externality. This is output increasing and welfare improving. If workers currently not working are reflected in the union's objective function, a union could even achieve the social optimum.

The question arises why these activities are mainly performed by government agencies in developed countries today. In our view, society as a whole has internalized the importance of understanding all aspects of new technologies. Early worker movements were not only successful in convincing companies and governments about health implications of, then, new technologies. The entire process also led to an understanding that a government should
have offices that systematically inquire into the health effects of new technologies. Putting this to the extreme, the success of trade unions made them redundant in modern societies where governments are sufficiently insightful into this important aspect of working conditions. That this extreme view is not always appropriate and that trade unions continue to play an important role also in advanced countries is visible from the discussion about work-life balance and burn-out syndromes. Both discussions are - historically speaking - fairly new and the respective role of individual characteristics of workers and work conditions is still to be seen. It might be a good idea for unions to go and collect data on these issues.

References


24