

July 2012

New and Larger Costs of Monopoly and Tariffs*

James A. Schmitz, Jr.

Federal Reserve Bank of Minneapolis

ABSTRACT

Fifty-eight years ago, Harberger (1954) estimated that the costs of monopoly, which resulted from misallocation of resources *across* industries, were trivial. Others showed the same was true for tariffs. This research soon led to the consensus that monopoly costs are of little significance—a consensus that persists to this day.

This paper reports on a new literature that takes a different approach to the costs of monopoly. It examines the costs of monopoly and tariffs *within* industries. In particular, it examines the histories of industries in which a monopoly is destroyed (or tariffs greatly reduced) and the industry transitions quickly from monopoly to competition. If there are costs to monopoly and high tariffs within industries, we should be able to see these costs whittled away as the monopoly is destroyed.

In contrast to the prevailing consensus, this new research has identified significant costs of monopoly. Monopoly (and high tariffs) is shown to significantly lower productivity within establishments. It also leads to misallocation within industry: resources are transferred from high to low productivity establishments.

From these histories a common theme (or theory) emerges as to why monopoly is costly. When a monopoly is created, “rents” are created. Conflict emerges among shareholders, managers, and employees of the monopoly as they negotiate how to divide these rents. Mechanisms are set up to split the rents. These mechanisms are often means to reduce competition *among* members of the monopoly. Although the mechanisms divide rents, they also destroy them (by leading to low productivity and misallocation).

Keywords: Monopoly; Competition; X-inefficiency; Rent seeking

JEL classification: D2, F1, L0

*I thank Tasso Adamopoulos, John Asker, Doug Clement, Tom Holmes, Pete Klenow, Sam Kortum, David Lagakos, Justin Pierce, Todd Schoellman, and Arilton Teixeira. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

In standard economic theory, monopoly leads to a welfare loss. This loss stems from a misallocation of resources *across* industries: too few goods are produced by the monopolist, too many in other industries. Economic theory had long suggested that this welfare loss exacted high costs on the economy. But modern understanding took a turn when, in a landmark 1954 paper, Arnold Harberger analyzed the quantitative significance of monopoly costs in the United States. Were these costs as high as conventional economic theory suggested? The clear but surprising answer that Harberger provided was no.

Harberger estimated that, contrary to his expectation and to standard theory, the costs of monopoly were quite trivial. “We come to the conclusion that monopoly misallocations entail a welfare loss of no more than a thirteenth of a per cent of the national income. Or, in present values, no more than about \$1.40 per capita,” he wrote. “I must confess that I was amazed at this result. . . . Monopoly does not seem to affect aggregate welfare very seriously through its effect on resource allocation” (Harberger 1954, pp. 85, 86, 87).

Other economists extended Harberger’s work to estimate costs associated with tariffs, and here, too, the costs were trivial. A consensus quickly developed that Harberger’s conclusion was indeed valid.

Recently, a new literature has taken a different approach to understanding the costs of monopoly. Looking *within* industries, it examines the histories of industries in which a monopoly is destroyed and the industry transitions quickly from monopoly to competition, as well as the histories of industries that rapidly moved the opposite way, from competition to monopoly. If there are costs to monopoly, we should be able to see these costs whittled away as the monopoly is destroyed. Likewise, if an industry is monopolized, we should be able to see the costs created by comparing the industry before and after monopolization.

Several industries have been studied with this method, including transportation in the United States and U.S. manufacturers of sugar, iron ore, and cement. The historical records of these disparate industries show that there are costs of monopoly and tariffs within industries. In these industries, this new literature has shown that monopoly led, among other costs, to the following:

1. *Low productivity* at each factory. That is, for any given amount of inputs, monopoly

meant that less output was produced than under competition.

2. *Misallocation of resources between high and low productivity factories.* That is, monopoly led to resources (capital, labor, etc.) being transferred from productive factories to unproductive factories. Again, this misallocation occurs within an industry and is different from the misallocation that Harberger (1954) studied.

In sharp contrast to Harberger’s finding, these studies show that the welfare costs associated with monopoly and tariffs are not small. The consequence of both cases (1) and (2) above is that industry output could have been produced with fewer inputs. Output was produced with inputs X , when it could have been produced with X' , $X > X'$. One way to measure the loss, then, is to calculate the value of the “wasted” inputs. Regarding the costs due to (1) above, that is, low productivity, as monopoly was destroyed in each of these industries, productivity at each factory soared. Doubling of productivities in a few years was common. The value of the wasted inputs was as much as 20 percent to 30 percent of industry value added.

Although this literature began by examining histories of industries, a common theme (or theory) emerges from the histories as to why monopoly led to these costs. When a monopoly is created, “rents” are created.¹ Conflict emerges among shareholders, managers, and employees of the monopoly as they negotiate how to divide these rents among themselves—or more colloquially, how to “split the spoils.” Mechanisms are set up to split the rents. Although they divide rents, they also destroy them (by leading to low productivity and to misallocation).

As used in this paper, the term “monopoly” means more than the strict definition: an industry with a single producer. One industry mentioned later in the paper was a cartel for 40 years. Conflict over rents emerged between different groups in the cartel, firms, workers, and managers. In some industries, there were high tariffs (and other forms of protection). This high protection led to strong incentives among groups in the domestic industry to form monopolies. Firms attempted to collude, and workers formed industry-wide unions (i.e., monopolies). So, the statement that “tariffs led to large welfare losses” means that tariffs led

¹In this usage, “rent” is the difference between what a factor of production is *actually* paid and what it would *need* to be paid to remain in use; as such, it is a measure of that factor’s monopoly power.

to incentives to form monopolies, and then to actual monopolies, and then these monopolies led to large welfare losses.

The remainder of the paper proceeds as follows. A body of literature in the 1960s and 1970s argued that the costs of monopoly and tariffs were not trivial, arguing (in essence) that there were costs within industries. I briefly review this theoretical literature in Section 2 and discuss the reasons why the literature did little to dent the “Harberger consensus.” Following this, I discuss the historical studies that look at the collapse of monopoly. First I discuss how the monopolies emerged and how they were destroyed. Then I discuss the mechanisms that were used to split rents and why these mechanisms led to welfare losses.

In the industries studied, a variety of mechanisms were created to split the spoils. One type of mechanism consisted of rules to limit competition *among* members of the monopoly. Another mechanism was direct side payments from one group in the monopoly to another.

Two types of competition-reducing mechanisms were used. One was *production quotas* given to firms in a cartel (to limit competition between them). Another was *work rules* in union contracts (to limit competition among workers). Production quotas and work rules are seen as very much the same: they both limited competition and both led to low productivity in establishments and misallocation within industry.

2. Early Arguments for Within-Industry Costs of Monopoly

Although a consensus quickly developed in the late 1950s and early 1960s that the costs of monopoly and tariffs were trivial, skeptics developed new ideas as to why monopoly was costly to society. The motivation of these researchers sprang from sentiments expressed by Mundell (1962) (whom they all cited). Regarding the “studies purporting to demonstrate that the welfare loss due to monopoly is small . . . and that gains from trade and the welfare gains from tariff reduction are almost negligible,” Mundell argued, “unless there is a thorough theoretical re-examination of the validity of the tools on which these studies are founded . . . someone will inevitably draw the conclusion that economics has ceased to be important!” (Mundell 1962, p. 622).

Notable among the new ideas in the 1960s was Leibenstein’s (1966) argument that monopolists do not minimize costs. Or to put it a different way, he argued that monopolists

produce with lower productivity than a competitive industry. That is, to produce a given level of output, a monopolist needs more energy, capital, and other inputs than a competitive industry would need.

Leibenstein's idea that monopolists do not minimize costs is obviously an argument that there are monopoly costs within industries. His idea certainly struck a chord. His term for this supposed phenomenon—X-inefficiency—is well known today. But despite its appeal, Leibenstein's theory did not have much influence on the mainstream economic literature. First, as Stigler (1976) and others argued, Leibenstein really had no theory to explain *why* a monopolist would not minimize costs. Second, and more important, he did not provide strong evidence that this was the case.

Other economists developed a different argument about the source of monopoly costs, pointing out that monopolies typically arise from government restrictions. Governments often restrict an industry to just one or a few firms by setting production quotas. Since these quotas are valuable, firms will compete for them in the political arena. The resources expended—on lobbying and the like—to obtain quotas are a cost of monopoly. Tullock (1967), and soon thereafter Krueger (1974) and Posner (1975), developed this line of thought. These costs of lobbying to achieve limited quotas (or, more generally, monopoly) came to be known as “rent-seeking” costs.

Tullock's idea that firms will compete for monopoly grants, and that resources used in this way are a loss, can also be seen as an argument that there are monopoly costs within industries. This idea certainly struck a chord as well: a large rent-seeking literature continues to this day. The idea, however, did little to change the Harberger consensus. First, as a matter of theory, the argument was made—by Rogerson (1982) and Fisher (1985), among others—that the costs of rent seeking could be small. For example, if one firm, because of its privileged position, was likely to win a contest for the quotas, it is possible that no resources were expended to capture the monopoly. Firms that knew they were likely to lose this contest would not waste resources in futile attempts.

Second, and most important, a large literature that attempted to “find” these rent-seeking costs essentially came up empty. The expenditures on rent seeking that were found were trivial (compared to the large welfare losses that some versions of the theory predicted).

Tullock (1998) discusses the failure of the literature to find significant welfare losses from rent seeking.

This terminology that divides monopoly costs into those arising from across-industry distortions and those arising from within-industry distortions is a useful one. It does have potential drawbacks, though. It may suggest to some readers that the costs of monopoly discussed here are “static” costs. That does not have to be the case, however, since some of the historical studies show how monopoly retards innovation.

On this point, let me briefly mention that in the 1980s, a new literature on monopoly took an entirely different direction, arguing that there were important benefits to monopoly: it led to innovation. This “endogenous growth” theory assumed that if someone developed a successful innovation, he would be immediately imitated and would thereby lose any opportunity to profit from his innovation. Hence, for innovation to occur, the promise of a monopoly (through a patent, for example) was necessary—at least for a period of time.

Although this literature grew by leaps and bounds in the 1980s and 1990s, it ultimately did little to swing the profession to the idea that monopoly (and high tariffs) was beneficial. First, other theories claimed the opposite.² Second, and most important, no strong evidence was given to support the theory. Moreover, a major effort by Boldrin and Levine (2008) found little evidence of the supposed benefits of monopoly as a spur to innovation.

At the turn of the century in 2000, then, the view established by Harberger (1954) nearly 60 years earlier was still the dominant one. However, the historical studies discussed later call the Harberger consensus into question. At least in the industries studied thus far, monopoly and tariffs have led to significant welfare losses.

3. New Approach to Studying Within-Industry Costs of Monopoly

Before discussing the historical studies that make up the new approach, let me briefly discuss why, on a priori grounds, we should expect this historical approach to help along two dimensions.

²Holmes and Schmitz (1995) and Parente and Prescott (1999) argue that monopolists block other firms’ innovations. Holmes, Levine, and Schmitz (forthcoming) argue that since innovation is sometimes disruptive to sales, greater monopoly, which means greater opportunity costs of lost sales, leads to less innovation.

3.1. Challenging the View that Monopoly Costs Are Trivial

The consensus that the costs of monopoly are trivial may be challenged in two ways. First, one can find new and sizeable costs of monopoly that have not been recognized. Second, one can provide evidence for costs that have already been discussed but dismissed as unimportant. The historical approach is ideally suited for both jobs.

As for finding new costs, there are arguably few better approaches than being present at the destruction of monopoly, to record and measure how the industry changes as it becomes competitive. If there are sizeable costs, then we should see them eliminated as the monopoly is destroyed.

As for providing evidence for old claims, consider the claim that monopolists do not minimize costs. The degree of skepticism about this issue has always been very high because, after all, why would monopolists fail to minimize costs (to maximize profits and not leave money on the table)? If we are going to dissuade people from this view, then, the evidence must be very clear and very strong—the type of evidence that a historical approach can provide.

3.2. Developing New Ideas as to Why Monopoly Is Costly

Although the initial purpose of these studies is to document costs of monopoly, there is good reason to think they can help in the development of theory as well. Suppose costs of monopoly—for example, low productivity—are documented. An obvious question then arises: Why did the monopoly produce with low productivity? The historical approach is valuable in trying to resolve such questions as well because we need not theorize in a vacuum. Lots of details about the industries, as well as the proximate causes of low productivity, are available.

3.3. Limits to the Approach

The historical approach, of course, does have its limits. These studies look at particular industries, asking what happens as monopoly is destroyed (because a cartel is abolished or tariffs are greatly reduced, and so on). By looking at industries in isolation, the studies cannot look at costs like those Harberger (1954) studied, namely, a misallocation of resources *across* industries. Interestingly, a series of recent papers have revisited this question of the costs of

tariffs that result from misallocation across industries (or across *goods*, in the terminology of these papers) (see, e.g., Holmes, Hsu, and Lee 2012, and Arkolakis et al. 2012). Edmund, Midrigan, and Xu (2012) and Ossa (2012) argue that, contrary to accepted wisdom, these costs are not small.

4. Monopoly: Its Creation and Destruction

In this section, I introduce some of the industries that have been studied, discussing how monopolies were created in the industries and how they were destroyed. In the section that follows, I discuss the costs of these monopolies.

In this paper, I discuss four industries studied with this approach: U.S. sugar manufacturing, in particular, sugar manufacturing using sugar beets (Bridgman, Qi, and Schmitz 2009, 2012); the U.S. iron ore manufacturing industry (Galdón-Sánchez and Schmitz 2002 and Schmitz 2005); the U.S. cement manufacturing industry (Dunne, Klimek, and Schmitz 2010); and U.S. freight transportation by water in the 19th century (Holmes and Schmitz 2001). These papers can be consulted for details that are only sketched in this paper. Other industries will be briefly discussed and references given at that time.

When a monopoly is created, the government often has a hand in the process. This is the case in most of the industries studied, to greater or lesser degrees. In U.S. sugar manufacturing, the government played a central role in creating monopoly. During the Depression, sugar manufacturers were permitted, indeed encouraged, by U.S. law, to form a cartel.

Many U.S. cartels were created during the Depression (as part of the New Deal), but the New Deal sugar cartel survived much longer than most. For 40 years, from 1934 to 1974, the industry was repeatedly able to renew the U.S. laws that enabled it to operate as a cartel. Soaring world sugar prices in 1974 resulted in the cartel losing political support, and the laws permitting it to operate as a cartel were not renewed.

In order to describe the government's role in creating monopoly in the other industries, a useful approach is to first sketch a very simple model, one in the spirit of that in Holmes and Schmitz (1995). Consider an industry where transportation costs are large relative to production costs. Suppose domestic production costs are c_D and that initially the domestic price is $p_D = c_D$. Suppose foreign production costs are c_F and transport costs from the

foreign to the local market are τ . Hence, as long $c_F + \tau > p_D = c_D$, the local price could be pushed upward until $p_D^{new} = (c_F + \tau)$ without risking entry of foreign producers. If demand were relatively inelastic over this range of prices, there would be an incentive to push price up toward $c_F + \tau$.

The incentive is great in this “industry” because, by assumption, transport costs are large relative to production costs. In particular, the percentage increase in price is

$$(1) \quad \frac{p_D^{new} - p_D}{P_D} = \frac{(c_F + \tau) - c_D}{c_D} = \frac{c_F - c_D}{c_D} + \frac{\tau}{c_D},$$

so that if $c_D \approx c_F = c$, the increase is τ/c_D , which, by assumption, is large. If $c_F > c_D$, then, of course, the incentive is even greater. In the previous formula, τ could have represented a tariff. So, a very large tariff will lead to an incentive to increase price, just as a large transportation cost would.

If the term in (1) is large, then, as in Holmes and Schmitz (1995), we will assume groups will make investments to form monopolies. Firms will attempt to collude, and workers to form strong unions. Some groups may succeed. If later on protection is cut (so τ falls) or the local cost advantage is narrowed (or eliminated), the incentives to make these investments will fall and the monopolies will weaken (or disappear).

This simple abstraction is a good representation of both the iron ore and cement manufacturing industries. In the early 1950s, U.S. producers had cost advantages over foreign producers, that is, $c_F > c_D$, and the industries received significant protection, that is, $\tau > 0$ (where τ stands in for tariffs and transport costs).³ Groups invested in creating monopolies. At various times, firms in these industries were charged with trying to collude. The U.S. government investigated the industries for antitrust violations. It is unnecessary to enter the argument as to how effective collusion was; there is little doubt that very strong, industry-wide unions emerged in these industries. Although antitrust laws in the United States made firm collusion difficult, building monopoly unions was easier. Collective bargaining laws enacted by the U.S. government allowed unions to organize all the workers in an industry and not be bound by antitrust laws (see Meltzer 1963 and Winter 1963).

³Transport costs were large (relative to production costs) in both industries. Also, the iron ore industry received government protection because the U.S. steel industry did. In the cement industry, the U.S. government at various times ruled that foreign producers were dumping in local markets.

The monopolies in these industries—in particular, the strong monopoly unions—lasted for many decades in the post–World War II period. The monopoly unions were able to provide very high wages. For example, by the 1970s, cement workers were paid as much as U.S. autoworkers (who were the highest paid manufacturing workers). The unions also had very stringent work rules (as described later on).

In the 1980s, the monopolies in these industries weakened or were dissolved. The union in the cement industry dissolved. In the iron ore industry, the union did not disappear but lost much of its clout. For example, work rules became much less stringent, and plant managers had more control over how to structure plant operations.

Why the weakening of the monopolies? The term in (1) had been getting smaller over the 1960s and 1970s. In the 1980s, the term actually changed sign and became negative. Foreign producers were now threatening to enter local markets. Brazil offered to sell iron ore in Chicago and Cleveland, the heart of the U.S. market, at half the local price. Firms across the world offered to sell cement on the West Coast and Gulf of Mexico at half the U.S. prices. With the term (1) negative, groups in the industry were no longer investing in monopoly.

Why did the term (1) become negative, so that foreign firms could offer to sell at such discounts? There are two proximate reasons. First, transportation costs greatly decreased (relative to production costs) in the post–World War II period. This, by itself, would have meant a weakened incentive for continued investment in keeping monopoly. But also the cost advantage of U.S. producers decreased. This development was, of course, to be expected, as the whole purpose of creating strong unions was to increase wages (and hence costs). The monopolies also led to lower productivity, increasing costs further. But what was striking is that U.S. producers were at a cost disadvantage.

An obvious question is: Why did the unions (and other groups discussed later on) push wages so high and lower productivity to the point where foreign producers could offer such steep discounts? At least three possibilities come to mind. First, the groups realized that wage demands and work rules would lead to the demise of monopoly, but that this strategy was the best. Second, the groups realized that wage demands and work rules would lead to the possibility of foreign entry, but they expected more government protection than they were able to receive. Lots of calls for protection were made, and some protection was given, but

it was not enough. Third, perhaps the outcome (foreign entry) was not expected. Although I do not know which story best describes the events, the story itself is not important for the issue at hand. The main point is that there are significant costs of monopoly and tariffs.

Monopolies can arise on their own, of course, without the help of government policy. Consider U.S. freight transportation by water in the 19th century. Well before any collective bargaining laws were enacted in the United States, strong unions developed in the port of New Orleans in the 19th century. Groups in transportation had a strong incentive to form monopolies in the port in the 19th century, since much freight went through that port. Many groups, such as warehouse owners, riverboat pilots, and longshoremen, were thought to have formed strong monopolies. Evidence of strong monopolies is particularly clear for longshoremen.

The weakening of these monopolies in New Orleans resulted from the development of alternative transportation technologies. New technology—railroads—meant that the returns from these port monopolies were greatly diminished. Investments in sustaining the monopolies waned, and the monopolies were considerably weakened.

5. Monopoly: Splitting the Spoils (and Destroying Them as Well)

During the period when monopolies in these industries were strong, groups set up mechanisms to split rents. Here I discuss some of the mechanisms used and how they led to the destruction of rents—in particular, to low productivity and misallocation. When monopoly was weakened in these industries, the mechanisms were abandoned, leading to large productivity gains in establishments and to resources being reallocated from low to high productivity producers.

5.1. Mechanisms Limiting Competition

One mechanism used to split rents was competition-reducing rules. Here I discuss two types of mechanisms that were used: *quotas* and *work rules*.

5.1.1. Quotas

In the U.S. sugar industry, the New Deal cartel included factory owners, factory workers, farmers, farmworkers, and others. As the cartel was established, each of these groups sought to secure (for themselves) as large a share of rents as possible. A major mechanism to split

rents was quotas. In the cartel, firms were given quotas—the right to sell a certain amount of sugar each year.

Incumbent farmers also sought, and were successful in acquiring, quotas—the right to grow sugar beet crops on a given number of acres each year. Without these quotas for incumbent farmers, nothing stopped firms from moving the locations of their factories or even using different farmers in the same location. Just as firms in the cartel used firm quotas to limit competition, incumbent farmers wanted quotas to limit competition among themselves (and from other farmers). Without these quotas, there was no way to ensure that incumbent farmers would receive a share of the monopoly profits.

As is often the case, these quota rights (both those of the firms and those of the farmers) could not be sold.⁴ Although the allocation of quotas for acres in 1934 was “efficient,” over time there was a change in the comparative advantage of locations in manufacturing sugar. Hence, there emerged a significant misallocation of resources between factories, with low productivity factories producing too much sugar and high productivity factories too little.

As the cartel started in 1934, some of the most profitable or productive (measured as revenue per acre divided by costs per acre) areas to make sugar were in California and Colorado. But after a few decades, these areas were no longer high productivity areas. The opportunity cost of land in California and Colorado, and of the water used in making sugar in these areas, grew much faster than in other parts of the country, in particular, in Minnesota and North Dakota. By the 1960s, these latter states became the most productive areas in which to make sugar. However, given the mechanisms to split rents (i.e., the quotas), the industry could not increase production in these areas. Once the cartel ended in 1974 and the mechanisms to split rents were abandoned, the share of industry production in Minnesota and North Dakota grew rapidly (and declined rapidly in California and Colorado).

We can estimate the magnitude of the welfare loss due to these mechanisms to split rents (i.e., the quotas), that is, from the misallocation of resources within the industry. Recall the introduction to this paper, which mentioned that one way to measure welfare loss is to

⁴Why the limit on selling quotas? When the cartel was being proposed, there were, obviously, complaints from farmers in sugar beet and sugarcane areas who did not grow these crops before 1934 (and hence were being left out of the cartel). A reasonable conclusion is that the cartel limited the quota rights so as to limit these complaints (thereby increasing the likelihood that the cartel would be acceptable).

calculate the value of wasted inputs in producing industry output. In the 1960s, the industry was using land in California, which had high value (or opportunity cost), rather than land in North Dakota, which had much lower opportunity cost. The difference in opportunity costs is a measure of the wasted land input. Not only were the opportunity costs of land much lower in North Dakota, but the opportunity costs of many other inputs—for example, labor and the water used in growing sugar beets—were much lower as well. To calculate the welfare loss in, say, 1965, we imagine “moving” some of the quota allocation from California to North Dakota (keeping industry output fixed) and then calculating the value of the inputs that were wasted by producing in California. At this point, it is easier to estimate the value of the wasted inputs relative to industry profits (rather than relative to value added). The estimates indicate that the losses were roughly 20 percent to 30 percent of industry profits.

5.1.2. Work Rules

In the iron ore and cement industries, those who were in a position to gain from the large transportation costs into local markets, and the protection offered by tariffs, were the factory owners, factory workers, and even the local governments (e.g., townships) where factories were located. What mechanisms were used to acquire rents? Local townships placed significant taxes on the production of iron ore and cement. Workers formed very strong unions. Although claims of collusion within both industries have been made, these claims are harder to document than the taxes and union contracts that emerged in these industries.

A major mechanism to split rents was the work rules in union contracts. Among other things, work rules were a way to limit competition among workers. They were structured so that managers could not play workers off each other. Let me briefly discuss these features of work rules and their consequences.

Union contracts split the tasks in plants into groups or categories. Workers were then assigned to one of these groups or categories, that is, given the right to complete tasks in that category. Only the workers in this group could complete the tasks assigned to the group. Very often these distinctions among workers were arbitrary in that a worker in a particular category was able, but not allowed, to complete tasks in many other categories.

Consider an important example. Machine operators were given tasks and repair work-

ers other tasks. Machine operators were prohibited from assisting repair workers in their assigned tasks, even mundane tasks that required no repair expertise, such as getting supplies, holding tools, and so on.

In addition to a sharp distinction being made between the tasks of machine operators and the tasks of repair workers, there was also a sharp distinction among tasks assigned to repair workers. Repair workers were grouped into many classifications, as many as 30 in a plant.

Consider a couple of examples of how repair work was divided among groups of repair workers. These two work rules were found in cement plants:

When the Finish Grind Department is completely down for repairs, the Company will not use Repairmen assigned to the Clinker Handling Department on repairs in the Finish Grind Department.

In cases where repair work on Mobile equipment . . . is required at times when Mobile Department Mechanics are not scheduled to work, the Repair Foreman will first attempt to contact the Mobile Mechanics to perform the work on an overtime basis. Should all of the Mobile Mechanics refuse the overtime or be otherwise unavailable to report to work, a General Repair crew will be assigned to do the job in conformity with past practices as to the nature of the repair work involved.

Both of these rules show how work is divided among repair staff, as well as the arbitrary nature of task assignment, in the sense that repair staff that could fix problems are not allowed to do so.

These types of work rules dividing work among members of the union are most often called *job classification systems*. They are similar to the quotas discussed earlier. In particular, work rules are a way to limit competition between workers, just as quotas limited competition among farmers. They ensure that groups of workers receive a share of the monopoly profits. But they also destroy profit, as I now discuss.

What are the negative consequences of such rules? These work rules in the iron ore and cement industries lead not only to overstaffing but also to idle machinery. When, for

example, the Finish Grind Department is down, workers from other departments are not allowed to help restore its machines to operation. Hence, machines are down longer than necessary, and capital productivity suffers. But clearly, energy productivity suffers as well. Fuel is being burned in other parts of the plant, and electricity used, even as the disabled machines are idle and output is not produced. As a result, such rules lead to low total factor productivity.

What is the quantitative significance of work rules? In the 1980s, when the work rules in the iron ore and cement industries were made much less stringent, labor productivity doubled in a few years. Other productivities increased as well. If these increases in productivity can be tied in large part to the relaxing of work rules, then obviously these are big welfare gains. The papers cited at the beginning of Section 4 for iron ore and cement manufacturing argue that most of the productivity gains were due to the relaxing of work rules. “Proving” such a claim is difficult, but the papers, by looking at both direct evidence and indirect evidence, have marshalled much evidence that this was indeed true.

We can estimate the magnitude of this welfare loss due to these mechanisms to split rents (i.e., the work rules), that is, from the low productivity in establishments. Again, one way to measure the welfare loss is by the value of the wasted inputs. With these work rules, machines were down longer than necessary. The energy that was being consumed elsewhere in the plant when output was not produced was a wasted input. The value of this wasted energy was its opportunity cost per unit multiplied by its quantity. The opportunity cost was its price per unit.

Next consider labor. With these work rules, labor input was wasted. For example, a machine operator could not hold a tool for a repair person (who would need to bring in another repair person for such tasks). The value of this wasted input was the opportunity cost of the machine worker’s time multiplied by the amount of time involved.

How much time was wasted in these plants? And what was the opportunity cost of this time? When work rules were changed in these industries, labor productivity doubled in a few years. Half of the workers were able to produce the same output. Suppose we use N to denote employment and assume that the wasted input was $(0.5) \cdot N$. Next, let w_m denote the (monopoly) wage paid in the industry, and let \tilde{w} denote the opportunity cost of labor, with

$w_m > \tilde{w}$. Let DWL denote the deadweight loss (due to wasted labor), and VA the industry value added. Then industry deadweight loss (due to wasted labor) relative to industry value added, where I use that value added in these industries is roughly half of industry revenue and labor's share of revenue is roughly 25 percent:

$$\frac{DWL}{VA} \approx \frac{\tilde{w} \cdot (0.5) \cdot N}{(0.5) \cdot p \cdot Y} \approx \frac{\tilde{w}}{w_m} \frac{w_m \cdot N}{p \cdot Y} \approx \frac{\tilde{w}}{w_m} \cdot (0.25).$$

Hence, if we assume a fairly big markup in wages, say, 50 percent, that is, $w_m/\tilde{w} = 3/2$, we have $\tilde{w}/w_m = 2/3$, and DWL/VA is roughly 16–17 percent. Smaller markups would mean greater losses.

In addition to energy and labor, capital was also wasted, as work rules meant that disabled machinery took longer to repair than was necessary. In considering estimating the welfare loss due to wasted energy and wasted capital, I note two things. First, energy productivity and capital productivity both increased significantly with the loosening of work rules, but not to the extent that labor productivity increased. So, not as much was wasted. On the other hand, the price paid for these inputs was likely close to its opportunity cost. Using a deadweight loss for the wasted capital and energy of a few percentage points (possibly more) of value added, together with the wasted labor estimate of 16–17 percent of value added, gives an estimate of over 20 percent in total.⁵

As just discussed, work rules in iron ore and cement manufacturing led to low productivity in establishments. But work rules can also lead to misallocation within industry. In Section 6, I briefly discuss some other industries, arguing that work rules likely led to the same type of misallocation as quotas did in sugar—with low productivity plants producing too much output, high productivity plants too little.

Finally, let me discuss U.S. freight transportation by water in the 19th century. In the port of New Orleans in the 19th century, work rules also led to low productivity. Rather than focus on these work rules, let me describe another type of misallocation that arose in

⁵In the calculation of the welfare loss due to the wasted labor input, I made no imputation for the value of leisure that might have been enjoyed at the plants (under the work rule regime). For example, perhaps the machine operator prefers standing next to the repair person not holding tools rather than helping out and holding tools. But this brings up other issues: What happens in a general equilibrium model if a large fraction of the population is standing around at work? We are looking at these issues elsewhere.

the ports.⁶ A few different types of dockworkers could be found at the port. One group was composed of workers called the cotton “screwmen,” a highly skilled group who loaded (screwed) cotton bales into the holds of oceangoing ships destined for Europe. Longshoremen, relatively unskilled workers on the docks, made up the other group. Both groups had strong unions and both were highly paid, with the skilled screwmen earning more than the longshoremen. The screwmen’s union was able to ensure that its workers not only loaded cotton ships destined for Europe, but also unloaded the cotton from riverboats entering New Orleans (which was unskilled work). This was a misallocation: high productivity workers doing unskilled work.

When the railroads came to the South and provided competition to the port, these work rules were relaxed or dropped, though making quantitative estimates for this episode is not as easy, given less evidence than in the other industries mentioned earlier.

The historical studies have shown that monopoly and tariffs can lead to welfare losses within industry on the order of 20 percent. But what about the losses for the entire economy? How many other industries, and of what size, have such incurred losses because of monopoly? I briefly discuss these questions in Section 6.

5.2. Side Payments Between Groups

Rules to reduce competition (such as quotas and work rules) were an indirect means to split rents between groups. Direct means were also used, whereby some factories would send money to other factories. This was done, for example, in the sugar cartel. These side payments in sugar manufacturing were not lump sum but involved mechanisms that led to distortions. In particular, the side payments exacerbated the misallocation problem discussed earlier (of having production in California and not North Dakota).

⁶As examples of work rules lowering productivity, deals struck by unions in the port of New Orleans in the 19th century required that machinery not be used in loading or unloading ships, or only in limited ways. Capital productivity suffered. Another type of port work rule in both the 19th and 20th centuries concerned cargo handling. When longshoremen removed a pallet from the deck of a boat and moved it to the dock, a rule forbade the pallet to be placed directly on the truck or wagon. First, rules required that the longshoremen take the cargo off the ship’s pallet and place it on the dock. Then a trucker would take the cargo on the dock and reload it onto his own pallet. The same rules held for loading material onto boats. This type of work rule obviously reduced capital productivity as well. The rules meant that unloading and loading boats took much longer than necessary. Hence, boats stayed in port much longer than necessary.

6. Splitting the Spoils: Other Industries and Other Countries

In Sections 4 and 5, I discussed industries that have been studied in great detail. In this section, I argue that the mechanisms to split rents that we saw in Section 5 are prevalent throughout industry. Moreover, evidence is suggestive that these mechanisms have had negative consequences in other industries as well. However, we cannot be sure of their quantitative significance because no studies like those described in the previous section have been completed for these industries.

Many U.S. industries had significant market power after World War II, first by virtue of the devastation that many countries faced as a result of the war, and later because of government protection of U.S. manufacturing. Monopolies emerged; in particular, the postwar years saw the emergence of industry-wide unions, such as in the auto, steel, paper, tire, airplane, and chemical industries, to name a few.

What mechanisms were used to split rents? The job classification systems discussed earlier are prevalent throughout manufacturing (though for the most part are less stringent today than a few decades ago). Some observers of these industries hold the view that work rules led to low productivity in plants.⁷

Just as they did in the cement and iron ore industries, stringent work rules likely led to low productivity in establishments in many manufacturing industries. In some of these manufacturing industries, stringent work rules led to other types of distortions and losses (which were not seen in the cement and iron ore industries). Here I discuss two other types of distortions. First, as I suggested earlier, work rules in these industries likely led to misallocation—of resources being transferred from high to low productivity plants. Consider the auto industry. For many decades after World War II, this industry was concentrated in Michigan, in particular, near Detroit. By concentrating this way, the industry gained from agglomeration economies. But over the last few decades, the industry has been moving from Michigan. The widely held argument is that the industry began to move South to avoid unions, with their attendant high wages and work rules. By so doing, the industry escaped

⁷See, for example, Hoerr (1988), who discusses the role of work rules in the U.S. steel industry collapse of the 1980s, and Simberg (2008), who discusses the recent auto industry crisis, lamenting the focus on high wages and not, in his view, the real culprit: work rules and low productivity.

the unions but also lost agglomeration benefits. Hence, the monopoly led to misallocation in industry: plants were moved from high agglomeration (high productivity) areas to low productivity areas.

Second, high wages (and stringent work rules) have likely led to another type of misallocation in industries: a change of technology (in order to escape the wages and work rules). In the U.S. steel industry, unions were very strong in large, integrated mills. This meant that the cost of labor per ton was much higher in the United States than in Japan and European countries (see, e.g., Miller 1984). This is likely one reason why minimill production, which was not typically unionized, grew so rapidly in the United States as compared to other developed countries. A similar phenomenon occurs in airlines. When airlines use jets that are under a given size (in the number of seats), they typically have more flexibility regarding wages and work rules. These regional jets are now used extensively, even on long flights.

A similar phenomenon, that is, monopolists splitting (and destroying rents), occurs in other countries as well. In Britain, job classification systems are also widespread. In that country, these work rules are called job demarcation rules—probably a better term for their purpose. In the 1960s, a Royal Commission (*Royal Commission on Trade Unions and Employer's Associations* 1968), known as the Donovan Commission, was established to study the causes of Britain's poor productivity performance, in particular, the idea that poor industrial relations was at fault. The commission found that work practices and work rules were indeed likely to be a significant cause of poor performance and emphasized the significance of demarcation rules in reducing productivity. Seidmann (1995) provides a brief discussion of the Donovan Commission findings, as well as some references to the academic literature on the subject. He also develops a model of demarcation rules.

In France, demarcation rules are also used. A famous study of industrial relations at a large French monopoly (Crozier 1964) discusses how the very poor utilization of machinery and equipment was due, as in the iron ore and cement industries, to rules that hampered the efficient repair of machines. “The problems people discuss most [regarding poor utilization] are machine stoppages and seniority rights. These two problems are linked, in as much as machine stoppages entail transfers that have to be handled according to seniority rights” (p. 67).

I finish this section by discussing a back-of-the-envelope calculation for the within-industry costs of monopoly and tariffs for the United States. In the industries studied thus far, within-industry welfare losses from monopolies and tariffs were on the order of 20 percent of industry value added. In order to turn this number into an aggregate U.S. estimate, we need to know what other industries could reasonably be considered to have such welfare losses. The point of this exercise is to make a preliminary stab at the question, are these welfare losses similar in magnitude to Harberger's losses (0.1 percent of value added), or can we conclude that they may well be significantly larger?

Industries that are well known to have strong unions and rigid work rules include mining, utilities, construction, transportation (in particular, airlines and railroads), and parts of manufacturing, in particular, durable manufacturing (steel, airplanes, autos, etc.). Very detailed job classification systems were used throughout these industries.

Let us consider making the welfare loss calculation for the late 1970s, when these unions were perhaps at their greatest strength. In the 1980s, competition in many of these industries soared as some industries were deregulated and others faced very strong foreign competition. As a consequence, work rules were relaxed. A reasonable assumption is that the work rules in these industries in the 1970s had similar impacts on productivity as they did in the industries discussed in detail earlier. Suppose we assume that welfare losses were on the order of 20 percent of value added in these industries. We next need a measure of the size of these industries. Yuskavage and Fahim-Nadar (2005) report that industry value added in 1977, as a share of GDP, was 2.1 percent in mining, 2.3 percent in utilities, 4.6 percent in construction, 13.1 percent in durable manufacturing, and 3.8 percent in transportation, or just over 25 percent in total. This would then mean that welfare losses amounted to roughly 5 percent of GDP. Again, this calculation is obviously extremely crude, but it does suggest that the losses may well be orders of magnitude larger than Harberger's losses.⁸

⁸Although work rules in the industrial industries discussed earlier have weakened considerably since the 1970s, in other industries, such as education, stringent work rules have grown. Moreover, occupational licensing has grown dramatically in importance, and its impacts on welfare may be important (see Kleiner and Krueger, forthcoming).

7. Costs of Monopoly: Summary and Observations

Research on the theoretical and quantitative significance of monopoly costs has evolved considerably since the mid-1950s. Prior to Arnold Harberger's (1954) influential paper, the prevailing view among economists was that the costs of monopoly, which resulted from misallocation of resources across industries, were considerable. Harberger's quantitative analysis suggested, in contrast, that in the United States, the costs were actually quite insignificant. This view soon became the dominant consensus among economists. Subsequent research in the 1960s and 1970s sought to establish a convincing counterargument, but was unsuccessful in overturning the prevailing concept of negligible costs.

This paper reviews a new stream of research that uses a different approach to analyzing the costs of monopoly. It examines the costs of monopoly and tariffs *within* industries. In particular, it examines the histories of industries in which a monopoly is destroyed (or tariffs greatly reduced) and the industry transitions quickly from monopoly to competition. Over considerable time spans and a wide range of industries, this research finds that monopoly exacts high costs in two ways: (1) through misallocation of economic resources between high and low productivity factories, and (2) by decreased productivity at each factory. The historical studies call the Harberger consensus into question. At least in the industries studied thus far, monopoly and tariffs have led to significant welfare losses.

A common thread runs through these histories, one that suggests a theory. When a monopoly is created, rents are generated. But the distribution of these rents—splitting the spoils—causes conflict among shareholders, managers, and employees of the monopoly. These parties devise mechanisms to split the spoils, but the mechanisms often lead, paradoxically, to the destruction of rents.

The historical studies of industries in some sense substantiate ideas developed in the 1960s and 1970s. First, they provide evidence for Leibenstein's (1966) view that monopolists do not minimize costs.

Second, the theory that emerges to explain the costs of monopoly is similar to Tullock's (1967) theory and the rent-seeking literature. It is in the conflict over rents that rents are destroyed. However, it is not conflict *among* lawyers of different firms that leads to the costs, but rather conflicts *within* monopoly firms. In evaluating the costs of monopoly, then, we

should not be looking at lobbying expenditures, but rather at what happens within the firms and organizations that receive protection. We should examine, for example, what happens to the productivity of organizations. In the studies thus far, protected organizations were found to have suffered significant productivity losses.

As for future economic research, a key question is to understand why mechanisms (such as work rules) are used to split rents when they also self-destructively wipe out rents. There is, as I mentioned, a large literature on models of rent seeking. These models typically involve games where players choose how much to spend to capture a prize. Rents are destroyed while fighting for the prize (see, e.g., Abbink et al. 2010 for a recent discussion). Although this literature has insights to offer, the big question remains about why members of the monopoly cannot structure contracts that avoid such large wasted resources. What are the fundamental reasons for this failure? Differences in information? The inability of parties to commit to future actions? Such reasons may well be why mechanisms to split rents also destroy them (see, e.g., McAfee and McMillan 1992 and Asker 2010 for a discussion in the context of auction markets).

References

- Abbink, Klaus, Jordi Brandts, Benedikt Herrmann, and Henrik Orzen. 2010. “Intergroup Conflict and Intra-Group Punishment in an Experimental Contest Game.” *American Economic Review* 100(1): 420–47.
- Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2012. “The Elusive Pro-Competitive Effects of Trade.” Working Paper. MIT.
- Asker, John. 2010. “A Study of the Internal Organization of a Bidding Cartel.” *American Economic Review* 100(3): 724–62.
- Boldrin, Michele, and David K. Levine. 2008. *Against Intellectual Monopoly*. Cambridge: Cambridge University Press.
- Bridgman, Benjamin, Shi Qi, and James A. Schmitz, Jr. 2009. “The Economic Performance of Cartels: Evidence from the New Deal U.S. Sugar Manufacturing Cartel, 1934–74.” Research Department Staff Report 437. Federal Reserve Bank of Minneapolis.
- . 2012. “A Cost of Monopoly: Misallocation of Resources from High to Low Productivity Plants.” Manuscript.
- Crozier, Michel. 1964. *The Bureaucratic Phenomenon*. Chicago: University of Chicago Press.
- Dunne, Timothy, Shawn Klimek, and James A. Schmitz, Jr. 2010. “Does Competition Spur Productivity? Evidence from the Post WWII U.S. Cement Industry.” Working Paper.
- Edmund, Chris, Virgiliu Midrigan, and Daniel Xu. 2012. “Competition, Markups, and the Gains from International Trade.” Working Paper 18041. National Bureau of Economic Research.
- Fisher, Franklin M. 1985. “Comment: The Social Costs of Monopoly and Regulation: Posner Reconsidered.” *Journal of Political Economy* 93(2): 410–16.
- Galdón-Sánchez, José E., and James A. Schmitz, Jr. 2002. “Competitive Pressure and Labor Productivity: World Iron-Ore Markets in the 1980s.” *American Economic Review* 92(4): 1222–35.
- Harberger, Arnold C. 1954. “Monopoly and Resource Allocation.” *American Economic Review* 44(2): 77–87.
- Hoerr, John P. 1988. *And the Wolf Finally Came: The Decline of the American Steel Industry*. Pittsburgh: University of Pittsburgh Press.

- Holmes, Thomas J., Wen-Tai Hsu, and Sanghoon Lee. 2012. "Allocative Efficiency, Mark-ups, and the Welfare Gains from Trade." Working Paper. Department of Economics, University of Minnesota.
- Holmes, Thomas J., David Levine, and James A. Schmitz, Jr. Forthcoming. "Monopoly and the Incentive to Innovate When Adoption Involves Switchover Disruptions." *American Economic Journal: Microeconomics*.
- Holmes, Thomas J., and James A. Schmitz, Jr. 1995. "Resistance to New Technology and Trade Between Areas." *Federal Reserve Bank of Minneapolis Quarterly Review* 19(1): 2–17.
- . 2001. "Competition at Work: Railroads vs. Monopoly in the U.S. Shipping Industry." *Federal Reserve Bank of Minneapolis Quarterly Review* 25(2): 3–29.
- Kleiner, Morris M., and Alan B. Krueger. Forthcoming. "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market." *Journal of Labor Economics*.
- Krueger, Anne O. 1974. "The Political Economy of the Rent-Seeking Society." *American Economic Review* 64(3): 291–303.
- Leibenstein, Harvey. 1966. "Allocative Efficiency vs. 'X-Efficiency'." *American Economic Review* 56(3): 392–415.
- McAfee, R. Preston, and John McMillan. 1992. "Bidding Rings." *American Economic Review* 82(3): 579–99.
- Meltzer, Bernard D. 1963. "Labor Unions, Collective Bargaining, and the Antitrust Laws." *Journal of Law and Economics* 6(Oct.): 152–223.
- Miller, Jack Robert. 1984. "Steel Minimills." *Scientific American* 250(5): 32–39.
- Mundell, Robert A. 1962. "Review of L. H. Janssen, *Free Trade, Protection and Customs Union*." *American Economic Review* 52(3): 621–22.
- Ossa, Ralph. 2012. "Why Trade Matters After All." Working Paper 18113. National Bureau of Economic Research.
- Parente, Stephen L., and Edward C. Prescott. 1999. "Monopoly Rights: A Barrier to Riches." *American Economic Review* 89(5): 1216–33.
- Posner, Richard A. 1975. "The Social Costs of Monopoly and Regulation." *Journal of Political Economy* 83(4): 807–28.

- Rogerson, William P. 1982. "The Social Costs of Monopoly and Regulation: A Game-Theoretic Analysis." *Bell Journal of Economics* 13(2): 391–401.
- Royal Commission on Trade Unions and Employers' Associations, 1965–1968: Report.* 1968. London: HMSO.
- Schmitz, James A., Jr. 2005. "What Determines Productivity? Lessons from the Dramatic Recovery of the U.S. and Canadian Iron Ore Industries Following Their Early 1980s Crisis." *Journal of Political Economy* 113(3): 582–625.
- Seidmann, Daniel J. 1995. "Organizational Change and Hysteresis in British Manufacturing Industry: A Perspective from the Donovan Commission Report." *Economica* 62(248): 507–20.
- Simberg, Rand. 2008. "Detroit's Downturn: It's the Productivity, Stupid." PJ Media. <http://pjmedia.com/blog/detroits-downturn-its-the-productivity-stupid/>.
- Stigler, George J. 1976. "The Xistence of X-Efficiency." *American Economic Review* 66(1): 213–16.
- Tullock, Gordon. 1967. "The Welfare Costs of Tariffs, Monopolies, and Theft." *Western Economic Journal* 5(3): 224–32.
- . 1998. "Which Rectangle?" *Public Choice* 96(3–4): 405–10.
- Winter, Ralph K., Jr. 1963. "Collective Bargaining and Competition: The Application of Antitrust Standards to Union Activities. *Faculty Scholarship Series.* Paper 2176. http://digitalcommons.law.yale.edu/fss_papers/2176.
- Yuskavage, Robert E., and Mahnaz Fahim-Nadar. 2005. "Gross Domestic Product by Industry, 1947–86: New Estimates Based on the North American Industry Classification System." *Survey of Current Businesses.* U.S. Department of Commerce, Bureau of Economic Analysis.