

Chapter Seven

TORRENS AND THE TERMS OF TRADE ARGUMENT

DESPITE SHARING with David Ricardo the credit for developing the concept of comparative advantage, Robert Torrens remains a relatively neglected member of the classical school of economists. A cogent and forceful advocate for free trade, Torrens also developed the most generally valid argument for tariffs. He described conditions under which a tariff could benefit a country by making the ratio at which it exchanged its products with the rest of the world—the terms of trade, or the purchasing power of a country's exports in terms of the imports it can procure—more advantageous. His strident and controversial criticism of unilateral free trade and advocacy of reciprocity in the 1840s triggered a sharp debate among economists. To argue against free trade was heresy in the minds of most economists of the day, and in crossing those bounds Torrens became a pariah for almost a century.¹ But the terms of trade argument for protection is the hardest to refute on theoretical grounds and remains the most durable and important exception to free trade ever conceived.

. . .

Torrens's early forays into economics were decidedly in favor of free trade. In his 1808 tract *The Economists Refuted*, he introduced the term "territorial division of labour" in rejecting the physiocratic notion that agriculture alone, and not international trade, contributes to wealth. In 1815, as discussed in chapter 6, he attacked protection to agriculture, stating that England's superiority in manufacturing was so great that it could import corn with advantage even if its own land was more productive than land in other countries. By this statement Torrens clearly anticipated the concept of comparative advantage as expounded by Ricardo in his *Principles of Political Economy*. Torrens also spoke out strongly in favor of free trade and dismissed any arguments for reciprocity. To the frequent contention

¹ The original *Palgrave Dictionary of Economics* dismissed Torrens's work as "devoid of permanent merit." Decades later the *New Palgrave* said that Torrens, "if not in the top rank of the classical economists, or in the class for example of Ricardo, Senior, or John Stuart Mill, certainly was of the second rank and was the equal of, or even above, James Mill and McCulloch in terms of originality, theoretical reasoning, and range of economic topics that he considered." See R.H.I. Palgrave (1913, 3: 550) and B. A. Corry (1987, 4: 659). For a masterful overview of Torrens's work, see Lionel Robbins (1958), especially chapter 7.

that it is “highly inexpedient in any one country to abandon the restrictive system [of import protection] while her neighbours continued to enforce it,” Torrens (1821, 268ff) responded that “nothing can be more erroneous or absurd.” A “mistaken policy on the part of France could furnish no conceivable reason why England should imitate the absurd example,” he wrote, adding that foreign protection did not force a free trade country to pay any financial cost or tribute, something he would explicitly contend in later work.

While Torrens and the classical economists were united in believing that the free exchange of goods between countries was mutually beneficial, they lacked insight into what determined the ratio at which goods would be exchanged between countries. Early expositions of comparative costs examples presumed a ratio of exchange exactly between the autarky cost ratios, allowing both countries to share equally the gains from exchange.² As Torrens (1821, 260) put it, “The advantages of foreign trade are reciprocal, and equally divided between nations carrying it on.” But the desirability of sacrificing a smaller rather than a larger bundle of goods through exports in exchange for a given quantity of imports (implying a high price of exports relative to imports) was also clearly recognized, even if there was little inkling about precisely how the terms of trade were established. Around 1820, Ricardo (1951, 2: 146) wrote that “it is undoubtedly true that if a country is to pay a certain money price for foreign necessities and conveniences, it is for its interest to sell the commodities at a higher, rather than at a lower price; it is desirable that for a given quantity of its own commodity, it should obtain a large rather than a small quantity of foreign commodities in return, but in what way a nation can so regulate its affairs as to accomplish this by any means which it is in its power to adopt, I am totally at a loss to conceive.”³

Despite the only rudimentary understanding of the determinants of the terms of trade, Torrens gradually came to discover that a country could shift the terms of trade in its favor by imposing a tariff. He used this theory to argue that tariffs should not be reduced unilaterally, as this would adversely affect the terms of trade, but in conjunction with other countries acting similarly under a policy of reciprocity. The groundwork for Torrens’s theory was developed in a series of letters to the *Bolton Chronicle* in

² On early errors in determining how countries divided the gains from trade, see William Thweatt (1987).

³ Ricardo (1951, 4: 71) staunchly supported free trade on a unilateral basis: “If foreign nations are not sufficiently enlightened to adopt this liberal system, and should continue their prohibitions and excessive duties on the importation of our commodities and manufactures, let England set them a good example by benefiting herself; and instead of meeting their prohibitions by similar exclusions, let her get rid, as soon as she can, of every vestige of so absurd and hurtful a policy.”

1832 as part of his effort to get elected to Parliament.⁴ Drawing on earlier ideas of Ricardo and Nassau Senior regarding the international distribution of precious metals, Torrens argued that tariffs could affect the movement of precious metals across countries. In particular, a country imposing a tariff would initially obtain a trade surplus, draw to itself a greater proportion of the world's precious metals, and thereby raise domestic prices, wages, and profits, and increase the purchasing power of its labor in terms of gold.

Holding his doctrines as immediately applicable to British circumstances, Torrens (1833, 6) maintained that the prosperity of the country depended on a trade policy based not on free trade, but on reciprocity. The guiding principle of commercial policy should be "to lower the duties of customs upon the importation of goods produced in countries which consent to receive British goods upon terms equally favourable, and to prohibit, or to lay heavy duties upon, the importation of all goods, not consisting of first necessities, produced in countries which prohibit, or lay heavy duties upon, British goods." Torrens accused the British government of departing from these principles, thereby reducing the price of British goods in foreign markets and undermining the country's superiority in manufacturing.

Torrens's initial, incomplete formulation neglected to stress the gain from improved terms of trade (or the international purchasing power of a country's exports) only to emphasize the deflationary consequences of unilateral free trade on domestic prices. Yet his heresy was evident and Torrens's ideas met broad resistance. Perronet Thompson (1833*a*) failed to understand how Torrens could argue that the nation as a whole could be worse off as the result of the free and profitable private trade of merchants. He further maintained that the export of specie was equivalent to the export of any other profitable commodity. Torrens (1833, 58) dismissed Thompson as having "no conception of the real question at issue."⁵ J. L. Mallet's diary entry gives an indication of the suspicion and hostility with which Torrens's views were received in a Political Economy Club meeting in 1835: "The first question discussed was a question of Torrens, which was unanimously voted to turn upon an impossible case. He claimed the right to discuss any abstract proposition with a view to the establishing of principle, but it was over-ruled in the present case which did not go to *establish* but to *disturb* a principle, that of Free Trade upon grounds altogether hypothetical" (Political Economy Club [1921, 270]).

⁴ These letters were collected and republished in Torrens (1833).

⁵ The meek reply is in Thompson (1933*b*). This exchange took a nasty turn. Torrens (1833, 57) called Thompson's work "correct, where not original, and where original, not correct," to which Thompson (1833*b*, 423) excoriated Torrens's "dishonest and in fact legally punishable action" of attacking his work in that way!

Torrens developed his views considerably more in a series of pamphlets addressed to leading political figures during the tariff debates of the early 1840s and later collected in a book entitled *The Budget: On Commercial and Colonial Policy*, published in 1844. Torrens became the leading exponent of the view that a unilateral tariff reduction would be detrimental to Britain's national welfare. His analysis hinged on two Ricardian concepts. First, international demand, and not costs of production alone, plays a role in determining the terms of trade.⁶ Second, commercial policies affect the international distribution of precious metals through the price-specie-flow mechanism.

From these precepts, Torrens (1844, 28) argued that "when any particular country imposes import duties upon the productions of other countries, while those other countries continue to receive her products duty free, then such particular country draws to herself a larger proportion of the precious metals, maintains a higher range of general prices than her neighbours, and obtains, in exchange for the produce of a given quantity of her labour, the produce of a greater quantity of foreign labour." This last statement, regarding the quantity of goods a country's labor could procure on the world market, emphasized the efficiency of trade restrictions in possibly procuring a large quantity of goods. The neighboring countries could recover these precious metals, he wrote, by retaliatory tariffs that would restore the previous exchange ratio.

Torrens then introduced a numerical example of tariffs and trade between Cuba (representing the rest of the world) and Britain to illustrate the proposition. If Cuba imposes tariffs from a situation of perfectly free trade, Britain will initially find itself importing the same value of goods from Cuba but exporting less. The concomitant trade imbalance will be financed by a flow of specie from Britain to Cuba, thereby reducing British prices and raising Cuban prices. The volume of trade would also adjust to ensure balanced trade, with the volume of British exports becoming larger and the volume of Cuban exports smaller. In the end, a greater amount of British exports would be exchanged for a smaller amount of imports from Cuba, an inferior ratio of exchange from Britain's perspective. While Torrens's numerical example was produced under particular assumptions, such as constant nominal outlays on Cuban products by Britain (implying unit elastic demand), his proposition was later shown to be more general in nature. Torrens (1844, 36–37) concluded that his example proved that the "ultimate incidence of the import duty imposed upon Brit-

⁶ As Ricardo ([1817] 1951, 1: 133) put it, "The same rule which regulates the relative value of commodities in one country, does not regulate the relative value of the commodities exchanged between two or more countries." Montiford Longfield (1835, 99–101) introduced demand more explicitly as a determinant of the terms of international exchange, but his contribution was incomplete and failed to attract much notice.

ish goods would be upon British producers. The wealth of England would be decreased by the amount of the duty—the wealth of Cuba would be increased by its amount.” Yet Torrens believed that this terms of trade effect “would be the least portion of the evil inflicted upon England by the change which has been described,” and instead the deflationary consequences were paramount—with “national bankruptcy and revolution . . . the probable result.”

Torrens’s policy recommendations sparked a controversy among economists that even spilled over into parliamentary debates. He (1844, 47–48) insisted that the following practical rules of commercial policy are “direct and necessary corollaries” from the principles he had described:

First,—to adopt, with respect to all foreign powers, the principle of reciprocity.—Second,—To lower the import duties upon the goods produced in countries receiving British goods upon terms equally favourable.—Third,—To impose high or prohibitive duties upon goods, the produce of countries imposing high or prohibitive duties upon British goods.—Fourth,—To admit, duty free, all raw materials employed in the processes of reproduction.

These rules could hardly constitute the commercial policy recommendations of a cosmopolitan free trader. In fact, Torrens was acutely aware of the distinction between national and world welfare, for two paragraphs later he wrote that “unrestricted interchange of commodities between different countries, would increase the wealth of the world.” Implicit in this statement is the recognition that the terms of trade improvement for Britain would imply a corresponding terms of trade deterioration for other countries, inflicting losses on others that exceeded Britain’s gains as the volume of trade shrank. But as Britain’s national welfare was at stake, he scolded the government for having “deprived the country of the advantages which our manufacturing superiority would otherwise have secured” and having “lowered the prices of British goods in foreign markets” (62). In addition, reciprocity would “hold out to [foreign countries] a powerful inducement to act upon the principles of reciprocal freedom” and perhaps lead to worldwide free trade (65). In sum, Torrens believed that “reciprocity should be the universal rule” and that “the sound principle of commercial policy is, to oppose foreign tariffs by retaliatory duties, and to lower our import duties in favour of those countries which may consent to trade with us on terms of reciprocity” (50).

Many economists, outraged at what they thought were Torrens’s irresponsible views, dismissed as irrelevant his whole analysis. As one anonymous reviewer put it, Torrens’s analysis was “without even a shadow of foundation.”⁷ Nassau Senior’s (1843) lengthy critique made him the most

⁷ “Colonel Torrens on Free Trade” (1843, 2).

prominent (and the most eminent) opponent of Torrens. Unfortunately for his cause, Senior's essay in support of unilateral free trade was a weak and desultory response.⁸ Despite this, the essay scored some telling points. Senior first accused Torrens of rejuvenating the erroneous doctrines of mercantilism. He (1843, 12, 14) then said that Torrens, while highlighting the adverse terms-of-trade impact of reducing tariffs, ignored the costs entailed by trade restraints:

Torrens assumes, first, that a country can exclude foreign commodities without diminishing the efficiency of its own labour. . . . It is a great mistake to suppose that a country which rejects the territorial division of labour, suffers merely by the greater dearness of the commodities which it is forced to produce instead of importing them. It incurs a further, and in many instances greater, injury—in the general diminution of the efficiency of its own industry, occasioned by the misdirection of capital and the diminished division of labour.

Senior accepted but dismissed Torrens's example of a tariff's impact on the terms of trade: "We believe this to be true; but we believe it to be one of those barren truths from which no practical inferences can be drawn. . . . In short, when he [Torrens] seriously urges us to act as if his hypothesis represented the actual state of things, we utterly dissent from, and repudiate his doctrine" (36–37). He also chided Torrens for assuming that Britain was the innocent victim of foreign tariffs when its own trade barriers were substantial as well.

Accurately noting that Senior was evasive in dealing with the main body of his argument, Torrens (1844, 350–51) jumped on the admission that Torrens's doctrine was true: "Your utter dissent and repudiation are utterly inconsistent with the facts and principles which you have yourself established . . . the doctrine which you admit to be true." Senior's review was also subject to a devastating attack by an anonymous author who laid out Torrens's ideas better than Torrens himself did.⁹ The author lambasted Senior for accusing Torrens of rehabilitating mercantilism: movements of precious metals were introduced only to restore trade to barter, the author sharply reminded the reader, and tariffs brought national gain not from acquiring metals *per se* but from changing the effective productivity of British labor in acquiring goods through international trade. The author easily demonstrated how supply and demand, and not costs of production as Senior continued to maintain, regulated international values and how

⁸ This opinion was held even by those who shared Senior's disdain for Torrens's views. "Senior's article contains very many just remarks, but he did not take the trouble thoroughly to understand T[orrens] before he sat down to answer him, therefore he is incomplete and unsatisfactory," opined S. J. Lloyd (Lord Overstone), a leading monetary theorist of the time. See the letter to G. W. Norman (December 13, 1843) in Denis O'Brien (1971, 345).

⁹ "Reciprocal Free Trade" (1843).

Torrens's example could, in principle, be extended to many countries and many commodities.

Herman Merivale (1842, 2: 305–11), one of the most incisive critics to emerge, restated Torrens's example of Cuba in purely barter terms not only to focus on the essence of the tariff and terms of trade argument but to show that the result was robust even when ignoring monetary adjustments. Merivale then introduced a second country, Brazil, that could supply sugar to Britain at a slightly greater cost than Cuba. If Cuba placed duties on imports from Britain, thereby raising the relative price of its sugar, England could simply switch its source of supply to "the next cheapest country producing the same commodities as Cuba." In all, Britain would be hurt only in proportion to the gap between Cuba's original price and Brazil's price of sugar, the trade of Cuba would be ruined, and Brazil would be the real beneficiary of Cuba's tariff. By allowing competition among Britain's import suppliers, Merivale demonstrated that Torrens exaggerated the impact of foreign tariffs if not all other nations increased their tariffs.

Torrens (1844, 358) reluctantly conceded this point, admitting that if Merivale's "assumption bore any resemblance to actual circumstances, the Cuba tariff could have a very slender effect in altering the terms of international exchange to the disadvantage of England." Torrens was left to assert that his example of all foreign countries increasing their tariffs was more appropriate.

George Warde Norman, in a lengthy publication written around 1845 and privately published in 1860, granted only that Torrens's case was half true. The burden of imposing retaliatory or reciprocity tariffs would also fall in part on foreigners, but some of the burden would fall on Britain's producers and consumers. Even though the burden would probably be borne equally by both countries, Norman (1860, 36) argued, the lower volume of trade and diminution of the natural advantages of trade would ensure that any gain would be "almost nugatory." He also made the point that, in practice, retaliation would never take place on the grounds established by Torrens, who had excluded from the policy of reciprocity any tariffs on raw materials, which, Norman pointed out, comprised the overwhelming proportion of British imports.

Torrens's analysis survived the onslaught of other theorists. James Anthony Lawson (1843, 133–47) held that the distribution of precious metals between countries was governed only by labor productivity in the export sector and tried to come up with an arithmetic example of a Cuban tariff that improved Britain's terms of trade. Torrens (1844, lii–lvii) charged that the first was a direct contradiction of Ricardian doctrine and showed, along with an anonymous reviewer of Lawson's work, that the arithmetic example had a fatal inconsistency.¹⁰ Implicitly assuming that the terms of

¹⁰ "Professor Lawson's Lectures on Political Economy" (1844).

trade are fixed and given from the perspective of a country, J. R. McCulloch (1849, 166) thought such tariffs would be “futile” as the burden on import duties would always be on the country imposing such tariffs. But McCulloch fully granted Torrens’s case for a country with a monopoly power in the export good, in which case a country could affect the international price of its exports.

This debate continued until John Stuart Mill weighed in with qualified support for Torrens. Mill had anonymously reviewed one of Torrens’s pamphlets in 1843 and shared his concern that foreign tariffs would accelerate the decline of Britain’s position as workshop of the world. While not directly discussing the terms-of-trade effects of tariffs, Mill (1843, 85–86) called foreign tariffs “the real source of alarm” for stifling British exports and potentially reducing the high wages commanded by British labor relative to competitors. He urged Britain’s politicians to spread free trade to other countries.

But Mill’s brilliant work on the determination of the terms of trade clinched Torrens’s argument and essentially settled the debate among economists. In 1844 in his *Essays on Some Unsettled Questions of Political Economy*, Mill published “Of the Laws of Interchange between Nations; and the Distribution of the Gains of Commerce among the Countries of the Commercial World.” This essay, originally written in 1829–30 according to Mill, described how world supply and demand determine the terms of trade between countries. In his preface to the collection of essays, Mill (1844, v–vi) wrote that they had been published “under the impression that the controversies excited by Colonel Torrens’s *Budget* have again called the attention of political economists to the discussion of the abstract science. . . . It will be seen that opinions identical in principle to those promulgated by Colonel Torrens (there would probably be considerable difference as to the extent of their practical application) have been held by the writer for more than fifteen years.”

“Of the Laws of Interchange between Nations” is the famous essay that set out the theory of reciprocal demand as the determinant of the equilibrium terms of trade. And it was here that Mill (1844, 21) questioned “whether any country, by its own legislative policy, can engross to itself a larger share of the benefits of foreign commerce than would fall to it in the natural or spontaneous course of trade.” He answered affirmatively and explained more clearly than Torrens the advantages of trade taxes if foreign demand for a country’s exports is not perfectly elastic. In such a case, the imposition of a tariff would reduce both import and export volume, with the reduction in export supply driving up the price of those exports on world markets. With the higher price, a given amount of exports could acquire a greater amount of imports than previously was the case.

Mill cautioned that while import duties may be advantageous under such

circumstances, “the determining circumstances are of a nature so imperfectly ascertainable, that it must be almost impossible to decide with any certainty, even after the tax has been imposed, whether we have been gainers by it or losers” (25). Furthermore, because the tax could eliminate a country’s trade if foreigners could buy from other sources, he argued that “even on the most selfish principles, therefore, the benefits of such a tax is always extremely precarious.”¹¹

Mill also endorsed a qualified version of a reciprocity-based trade policy, distinguishing between a protecting duty, which encourages a particular branch of domestic industry by attracting labor and capital to its production, and a revenue duty, levied on those goods not produced at home. A “protecting duty can never be a cause of gain, but always and necessarily of loss, to the country imposing it,” but Mill stated that, with revenue duties,

considerations of reciprocity, which are quite unessential when the matter in debate is a protecting duty, are of material importance when the repeal of duties of this other description is discussed. A country cannot be expected to renounce the power of taxing foreigners, unless foreigners will in return practise towards itself the same forbearance. The only mode in which a country can save itself from being a loser by the duties imposed by other countries on its commodities, is to impose corresponding duties on theirs. (28–29)

Mill noted that a country could only improve its terms of trade at the expense of other trading countries, and their losses would exceed the gains of the tariff-imposing country. Therefore, “it is evidently the common interest of all nations that each of them should abstain from every measure by which the aggregate wealth of the commercial world would be diminished, although of this smaller sum total it might thereby be enabled to attract to itself a larger share.” But “until, by the common consent of nations, all restrictions upon trade are done away, a nation cannot be required to abolish those from which she derives a real advantage, without stipulating for an equivalent” (31–32). Still, Mill believed that import duties would bring harm to all countries, doubting that tariffs could be properly set to the advantage of any. He observed with concern the severely protectionist policies in France, the Netherlands, and the United States, arguing that such policies, “though chiefly injurious to the countries imposing them, have also been highly injurious to England” (38).

Thus Mill accepted (indeed originated, though Torrens popularized) the theoretical point that a tariff can improve a country’s terms of trade. Mill

¹¹ Mill set out his example in both barter and monetary terms. On the latter, Mill (1844, 40–41) noted what Torrens had stressed, that tariff reductions may have a deflationary effect, but added that this gave “rise, as a general fall of prices always does, to an appearance, though a temporary and fallacious one, of general distress.”

stressed that such tariffs were a negative sum game for the world as a whole (a point recognized but downplayed by Torrens, who was concerned about national gains). As a result, Mill exercised great restraint in drawing specific policy recommendations from the proposition. His (1843, 85) caveat was that Torrens, “as is not unusual with him, seems to us to overstate the importance and urgency of a portion of his doctrines in their application to the immediate circumstances of the country.” That Torrens was correct in theory was confirmed when the controversy prompted Mill to publish his previously written essay on the subject. But as to the policy recommendations that Torrens considered natural conclusions of his analysis, Mill hesitated endorsing them and other economists rejected them entirely.

Torrens and Mill had developed a theory that consisted of two parts: first, under certain circumstances a tariff reduction could lead to a deterioration in the terms of trade (or, conversely, a tariff increase could improve the terms of trade); second, a country undertaking such a tariff reduction could conceivably suffer a net economic loss as a result. Mill and Torrens demonstrated the first proposition conclusively. The second proposition remained speculative: would the lower tariff bring about an adverse terms-of-trade effect sufficient to outweigh the gain from the expansion of trade resulting from a greater international division of labor? Torrens and Mill assumed that an improvement in the terms of trade by itself would necessarily imply greater economic wealth. But this ignored the equally important contribution of the volume of trade to the gains from trade, and hence their cost-benefit analysis was inadequate. As F. Y. Edgeworth (1894, 40) later noted, “Mill obscures the subject by taking as the measure of the gain of trade the alteration in the rate of exchange between exports and imports rather than the truer measure of advantage which the principles of consumers’ and producers’ rent afford”—or, in modern terms, consumer and producer surplus as a measure of economic welfare.

It was not until the late nineteenth century that Edgeworth, an economist and statistician at Oxford, demonstrated the validity of the second proposition on more rigorous analytical grounds. Edgeworth employed a construction called the offer curve, developed by Alfred Marshall in the early 1870s to illustrate various combinations of export and import volume that a country was willing to exchange with others, to represent graphically Mill’s reciprocal demand theory.¹² Edgeworth (1894, 433ff) ingeniously combined these offer curves with a graphical representation of a country’s aggregate economic welfare (through utility indifference curves, depicting the country’s preferences across various goods). Edgeworth was thereby able to sketch out a proof that, if the foreign (rest of the world) offer curve

¹² Marshall privately distributed a manuscript describing these curves in 1879 but did not initially apply this apparatus to commercial policy. He later reprinted them in part and discussed tariff policy in Appendix J of Marshall (1923).

was anything but perfectly elastic (in which case the terms of trade would be fixed by the world market), there was the potential for tariffs to improve the terms of trade and increase national welfare. The “optimal” tariff became the term used to describe a tariff that exploited the terms of trade precisely to maximize national economic welfare. Edgeworth also confirmed the Mill-Torrens view that this gain arose entirely by imposing an even greater loss on other countries. Edgeworth appears not to have recognized the significance of his findings until later, but he secured for the terms of trade argument the final clinching verdict. Any lingering doubts about the theoretical possibility of achieving higher economic welfare under these conditions through tariffs was eliminated.

Edgeworth’s work later convinced Alfred Marshall, the great Cambridge University economist, to repudiate use of the analytical tools he himself had developed. “In recent years, I have gradually gone away from the fundamental hypothesis on which the curves are based,” Marshall (1925, 449) wrote to a correspondent in 1904. “They lead to the result that a great part of an import duty will probably fall on the export nation: and I have become convinced that, though the reasons which the old free-traders gave for the opinion that import duties are paid almost entirely by the consumer are wrong, yet their result is pretty well true.” Despite Marshall’s concern, his offer curve construction became a standard tool in the theoretical analysis of tariffs. Edgeworth’s analysis was later revived and elaborated by Nicholas Kaldor (1940) and others in the 1940s. This approach culminated in the early 1950s with the derivation by Harry Johnson (1950–51) of a precise mathematical formula for the “optimal” tariff based on the elasticity of the foreign offer curve.

Aside from these advanced analytical deliberations, another debate concerning Torrens’s and Mill’s theory focused on the closely related question of tariff incidence—or “who pays the tariff?”—when a country possesses market power in its export or import markets and does not take world prices as given. Some classical economists, such as J. R. McCulloch (1849, 166) as we have seen, implicitly assumed that a country could not influence the prices of its imports as determined by world market, and therefore concluded that domestic consumers bore the full burden of the tariff. Under this scenario, the idea that tariff revenue could be extracted from foreign suppliers “is wholly visionary, and that duties on imports are always paid by the importers and never by the exporters” Advocates of tariffs, of course, took the other extreme position and asserted that the burden of import tariffs could be shifted entirely onto foreigners through lower import prices.

Mill, appropriately, took a middle path. As a result of his theory of reciprocal demand, Mill (1844, 27) pointed out, “it may, therefore, be laid down as a principle, that a tax on imported commodities, when it really operates

as a tax, and not as a prohibition either total or partial, almost always falls in part upon the foreigners who consume our goods; and that this is a mode in which a nation may appropriate to itself, at the expense of foreigners, a larger share than would otherwise belong to it of the increase in the general productiveness of the labour and capital of the world, which results from the interchange of commodities among nations.”

But the state of knowledge about the question of tariff incidence remained uncertain through most of the nineteenth century. Henry Sidgwick (1883, 492–93) argued that “there is no theoretical means of determining” the impact of a tariff on domestic consumers or producers, but “unless foreign products are completely excluded by import duties, such duties will partly have the effect of levying a tribute on foreign producers, the amount and duration of which may in certain cases be considerable.” This arises either because the foreign costs of production (and hence its price) will fall as its export volume falls, or because profits could be extracted from foreign monopolists or cartels. On the export side, Edgeworth (1894, 42–43) reiterated Mill’s argument about how export restraints could be advantageous in raising export prices (if foreign demand for a country’s goods was somewhat inelastic), noting that “it is often stated with the unnecessary limitation that the home country must have an absolute monopoly of the exporting article,” whereas “that she should furnish a considerable portion of the total supply might suffice.” It was later pointed out that if domestic producers do not act as perfect competitors in the export market but recognize the market power that they collectively possess, these producers will collude to restrict their exports and hence mimic the optimal export tax, obviating the need for government intervention.

Yet outlining the possibility of such an outcome was quite different from actually proposing tariffs for this purpose. Joseph Shield Nicholson (1891, 465) summed up the view of many economists at the end of the nineteenth century in writing that

taxing the foreigner is very like “shearing the wolf.” It is quite true that *theoretically* under certain conditions one nation might obtain from other nations, either by export or by import duties, a considerable part of its revenue, but it is equally true that these conditions are extremely unlikely to arise; and even if they did arise, it is still more unlikely that the wisdom of statesmen would be equal to the task of taking advantage of them. It is important to observe that theoretical exceptions may be admitted whilst the practical application is denied, for no greater harm has been done to the spread of “Free Trade principles” in the broad sense of the terms than by the attempt to reduce them to a fictitious simplicity. To assert that every import-duty must *necessarily* fall on the home consumer is as false as to assert that every export-duty must necessarily fall on the foreign consumer; it is equally untrue to say that *necessarily* the import-duties fall on the foreign producer and the export-duties on the

home producer. As a matter of fact the incidence of export and import duties, especially when the indirect effects are considered, is the most complicated and difficult problems in economics.¹³

The “complicated and difficult problem” of determining tariff incidence and whether the tariff burden could be shifted onto foreigners soon gave way to an analytical breakthrough. Using Marshall’s supply and demand schedules to represent the market for an imported good in partial equilibrium, Charles F. Bickerdike (1906, 529ff) raised the question of whether “a country, by means of taxes, can get more favourable terms of exchange with foreigners in such a way as to leave a net advantage, after allowing for the disadvantages involved in turning production from its ‘natural’ course.” He answered affirmatively: “In pure theory advantage is always possible in normal circumstances from either import or export taxation when the taxes are small enough.” Provided only that the foreign export supply curve is upward sloping, some of the revenue generated by a tariff is extracted from the producer surplus of foreign suppliers. Furthermore, this is “a general possibility of advantage not confined to exceptional circumstances.” Bickerdike demonstrated with the simple Marshallian geometry that, for small or “incipient” taxes, a net gain results: the higher domestic price paid by consumers is largely a transfer to the government treasury, but the small deadweight loss of consumer surplus is exceeded by the gain from lower foreign prices over the entire volume of imports (and collected as tariff revenue) at the expense of foreign producers.

Bickerdike later extended these findings to derive a mathematical formula for the optimal tax based on export and import supply and demand elasticities. Under this formula, he (1907, 101) noted that “rather strong assumptions have to be made as to the elasticity of foreign supply and demand if the rate of the tax affording maximum advantage is to come below 10 per cent.” In evaluating this theory, Edgeworth (1908, 392) praised Bickerdike for his having “accomplished a wonderful feat” of saying “something new about protection” and reinforced the point that Bickerdike was correct as a matter of theory. But Edgeworth was deeply skeptical about the practical application of the finding. Conceding that Bickerdike’s result appeared to be quite general and only required information about supply and demand elasticities, Edgeworth (1908, 554) remarked that the theory seemed

to justify the imposition of small customs duties, say from 2½ to 5 per cent, on a great number of articles. The objection that industry is thereby directed into less advantageous channels is not admissible; for by the theory the disad-

¹³ Nicholson (1901, 306) later wrote that “certain exceptional conditions under which it is theoretically possible by the judicious manipulation of duties to extract a certain amount of revenue from the foreigner” although “practically they are of little importance.”

vantage in the way of production is overbalanced by the gain accruing to the Treasury. Abstracting the practical difficulties to which we are coming, on the platform of pure theory the Free Trader must abandon his hectoring tone with respect to the defence of a Protectionist tax on the ground that it is a little one.

But Edgeworth went on to mention “weighty objections” that gave the theory “limited application.” Aside from the frictions that many small taxes would entail in practice, the foremost obstacle was the threat of retaliation. Bickerdike’s novel and ingenious sketch may imply a strong ability of hitting the foreigner with taxes because it “may be practiced by a country which has no special [monopoly] advantages” in trade, “but it equally increases the power of the foreigner to hit back.” For these reasons, Edgeworth (1908, 555–56) continued,

the direct use of the theory is likely to be small. But it is to be feared that its abuse will be considerable. It affords to unscrupulous advocates of vulgar Protection a peculiarly specious pretext for introducing the thin edge of the fiscal wedge. Mr. Bickerdike may be compared to a scientist who, by a new analysis, has discovered that strychnine may be administered in small doses with prospect of advantage in one or two more cases than was previously known; the result of this discovery may be to render the drug more easily procurable by those whose intention, or at least whose practice, is not medicinal. . . . Let us admire the skill of the analyst, but label the subject of his investigation POISON.

Just as the foreign reciprocal demand approach had been resurrected in the 1940s, the partial equilibrium “elasticities” approach to the optimal tariff was also resurrected then by, among others, Richard Kahn (1947–48), who supported Bickerdike’s contention that the optimal tariff could be substantial based on what was believed about the magnitude of the elasticities.

Thus, although the terms of trade argument continued to undergo further refinements, few developments have seriously undermined its standing as a theoretically valid proposition. The most commonly proposed qualification is that foreign retaliation might undo the positive terms of trade effect of one country’s optimal tariff. Since a country could only improve its terms of trade at the expense of other countries, an attempt to do so might incite the other countries into responding in kind to preserve their own position. And if all countries use tariffs in seeking to improve their terms of trade, the result would simply be a contraction in the volume of trade without anyone succeeding in their effort. One conceptual qualification to this argument was provided by Harry Johnson’s (1953–54) finding that it was at least possible that, even after retaliation, one of the countries could still be better off than it would have been under free trade. Yet Johnson showed a possibility, not a likelihood, and the threat of foreign retaliation

has been an oft-invoked practical argument against the terms-of-trade motive for trade intervention.¹⁴

The conclusion to be drawn from the terms of trade controversy is not that free trade is undesirable, but that, under certain circumstances, *unilateral* free trade is undesirable. Therefore, to avoid a situation in which countries seek to gain at the expense of others by imposing optimal tariffs, thereby diminishing the worldwide gains from trade, countries may wish to have a commitment mechanism, an agreement, in which all agree to forgo the use of tariffs for this purpose. In this context, contractual multilateral free trade avoids some of the problems associated with countries seeking to manipulate their terms of trade. And, of course, the cosmopolitan case for free trade remains unaffected. As John Stuart Mill (1844, 44) put it, "If international morality, therefore, were rightly understood and acted upon, such taxes, as being contrary to the universal weal, would not exist."

. . .

After a brief but intense debate, Torrens's speculations that tariffs could conceivably be beneficial in improving the terms of trade gradually become orthodoxy under Mill, who demonstrated the soundness of these speculations in terms of its underlying economic analysis. Although inadvisable from a global perspective for shifting the gains from trade to one country at the greater loss of others, few developments have detracted from its validity as a theoretical proposition. Indeed, of all the economic arguments against free trade, the terms of trade argument appears to be the most robust and least subject to qualification or exception, and it remains the most widely acknowledged and generally accepted restriction to free trade admitted by economic theory.

¹⁴ Carlos Rodriguez (1974) showed that both countries would definitely be worse off if retaliation took place by means of import quotas instead of import tariffs.

Chapter Eight

MILL AND THE INFANT INDUSTRY ARGUMENT

JOHN STUART MILL, the preeminent figure in mid-nineteenth-century economics, wrote the *Principles of Political Economy* (1848), which became the standard economics treatise for several generations of students. Yet one paragraph of Mill's influential text inspired a lasting controversy and earned him the scorn of many of his contemporaries. In a brief passage, Mill gave his qualified endorsement to temporary protection for "infant industries," defined as industries that were not initially capable of surviving in the face of import competition but with time and experience could grow to compete successfully in world markets. Mill caused great consternation among economists by sanctioning this argument for protection, although he eventually (but quietly) recanted his endorsement of tariffs as the means of supporting such industries. Despite its somewhat vague theoretical formulation, the infant industry argument survived many criticisms and continues to occupy an uneasy place in the theory of commercial policy.

. . .

The infant industry argument, which is perhaps the oldest and longest-lived specific argument for protection, can be traced at least as far back as the Elizabethan period. Precise statement of the doctrine, involving temporary protection to establish a new industry, arises in the mercantilist period. Viner (1937, 71) finds a passage from 1645 arguing that monopoly privileges in a particular trade were no longer necessary because the domestic firm, having matured, was no longer in its infancy. More common were pleas for government support to assist fledgling manufacturers against foreign competition. Andrew Yarranton (1677, 62), for example, advocated

that the linen and iron manufactures may be so encouraged here by a public law, as that we may draw these trades solely to us, which now foreign nations receive the benefit of, there ought in the first place to be a tax or customs at least of four shillings in the pound put on all linen yarn, threads, tapes, and twines for cordage that shall be imported into England . . . and this law to continue and be for seven years. And by virtue of this tax or imposition, there will be such advantage given to the linen manufacture in its infancy, that thereby it will take deep rooting and get a good foundation.

William Wood (1718, 224–25) repeated verbatim the statement by Theodore Janssen (1713, 9) that “all wise nations are so fond of encouraging manufactures in their infancy, that they not only burden foreign manufactures of the like kind with high impositions, but often totally condemn and prohibit the consumption of them.” Arthur Dobbs (1729, 2: 65ff) argued similarly: “Upon the whole, premiums are only to be given to encourage manufactures or other improvements in their infancy, to usher them into the world, and to give an encouragement to begin a commerce abroad; and if after their improvement they can’t push their own way, by being wrought so cheap as to sell at par with others of the same kind, it is in vain to force it.”

The infant industry case was a part, although by no means a pervasive part, of the mercantilist desire to promote domestic employment and industry. These few early statements adequately convey the main thrust of the infant industry doctrine, that before new industries could compete successfully against established foreign rivals, government assistance was required to help them overcome certain start-up obstacles and grow to maturity. The appealing and intuitive metaphor of an “infant” lured most eighteenth-century economic writers into accepting the case for infant industry protection without much serious questioning. Even Adam Smith’s teacher, Francis Hutcheson (1753, 308), seemed to accept the doctrine when he wrote “all mechanick arts, either simpler, or more elegant, should be encouraged, lest our wealth be drained by our buying foreign manufactures.”

The infant industry argument was so widely accepted that writers made their mark by pointing out qualifications to it. Josiah Tucker (1758, 50–51), for example, cautioned that protection should be temporary or else the infant industry may never mature:

It is also easy to see, that such infant manufactures, or raw materials, as promise to become hereafter of general use and importance, ought to be reared and nursed during the weakness and difficulties of their infant state, by public encouragements and national premiums. But it doth by no means so clearly appear, that this nursing and supporting should be *continued forever*. On the contrary, it seems more natural to conclude, that after a reasonable course of years, attempts ought to be made to wean this commercial child by gentle degrees, and not to suffer it to contract a lazy habit of leaning continually on the leading strings . . . that trade is not worth the having, which never can be brought to support itself.

Others argued that the tariffs used to promote infant industries should not be set too high. In Malachy Postlethwayt’s (1757, 2: 397) view:

A manufacture, even in its infancy, seems not to have any reason to fear foreign competition, when the duties of entry are at 15 per cent; for the charges of

carriage, commission, and others, will come to 4 or 5 per cent more. If 18 or 20 per cent, besides the foreign manufacturer's profit, do not content the home manufacturer; one may readily conclude, either that such home manufacture wants to gain too much, or, that his undertaking is badly managed; or, in short, that there are such obstacles in the way at home, which must be removed before success can be expected.

On the other hand, James Steuart ([1767] 1966, 262–63) expounded the infant industry argument with great enthusiasm and little qualification: “The ruling principle, therefore, which ought to direct a statesman in promoting and improving the infant trade of his people, is to encourage the manufacturing of every branch of natural productions, by extending the home-consumption of them; by excluding all competition with strangers; by permitting the rise of profits, so far as to promote dexterity and emulation in invention and improvement; . . . and, until it can be exported to advantage, it may be exported with loss, at the expence of the public.” Steuart had no concern that protection would produce slothfulness rather than promote improvement, for “as long as the gates of a kingdom are kept shut, and that no foreign communication is permitted, large profits do little harm, and tend to promote dexterity and refinement.”

Although economic writers differed in the degree to which they qualified their support for infant industry policies, one is hard pressed to find anyone who actually disputed the basic argument prior to Adam Smith. It is a credit to his intellectual independence that Smith opposed government support for infant industries even though other Scottish philosophers (such as Lord Kames) endorsed it. Perhaps in response to unqualified statements like Steuart's, Smith reacted so strongly against the infant industry doctrine that, unlike his predecessors, he came close to denying that such protection could ever be beneficial:

By means of such regulations, indeed, a particular manufacture may sometimes be acquired sooner than it could have been otherwise, and after a certain time may be made at home as cheap or cheaper than in the foreign country. But though the industry of the society may be thus carried with advantage into a particular channel sooner than it could have been otherwise, it will by no means follow that the sum total, either of its industry, or of its revenue, can ever be augmented by any such regulation. The industry of the society can augment only in proportion as its capital augments, and its capital can augment only in proportion to what can be gradually saved out of its revenue. But the immediate effect of every such regulation is to diminish its revenue, and what diminishes its revenue, is certainly not very likely to augment its capital faster than it would have augmented of its own accord, had both capital and industry been left to find out their natural employments. Though for want of such regulations the society should never acquire the proposed manufacture, it would

not, upon that account, necessarily be the poorer in any one period of its duration. In every period of its duration its whole capital and industry might still have been employed, though upon different objects, in the manner that was most advantageous at the time. In every period its revenue might have been the greatest which its capital could afford, and both capital and revenue might have been augmented with the greatest possible rapidity. [*WN*, IV.ii.13–14]

To Smith, the underlying cause of the foreign industry's superiority was irrelevant: "Whether the advantage which one country has over another, be natural or acquired, is in this respect of no consequence. As long as the one country has those advantages, and the other wants them, it will always be more advantageous for the latter, rather to buy of the former than to make" [*WN*, IV.ii.15]. Indeed, Smith was so deeply skeptical of the infant industry argument for protection that he conceded virtually nothing to it. Just because a country could acquire an industry by means of such protection did not imply that it should do so, or that the country would be better off for having done so. And even if a country could eventually produce that industry's good at a lower price than foreign producers, the policy still might be disadvantageous: protection would distort resource allocation, reduce national income, and thereby shrink the pool of investible savings available for capital accumulation. Smith also implicitly stressed the intertemporal balancing of the policy's costs and benefits, that the short-run costs must be offset by some distinct, long-run benefit, a point that was to be ignored for decades. However, by taking a fairly static view of infant industries, Smith was vulnerable to critics who would complain that he failed to deal with the essentially dynamic issues involved.

Other classical economists followed Smith's lead on the infant industry argument, either treating it unfavorably or ignoring it altogether. Jean-Charles-Léonard Simonde de Sismondi (1815, 70) focused on the opportunity costs of shifting by artificial means scarce resources into favored sectors at the expense of other sectors: "It ought to be recollected that each merchant knows his own business better than the government can do; that the whole nation's productive power is limited; that in a given time, it has but a given number of hands, and a given quantity of capital; that by forcing it to enter upon a kind of work which it did not previously execute, we almost always at the same time force it to abandon a kind of work which it did execute: whilst the most probable result of such a change is the abandonment of a more lucrative manufacture for another which is less so, and which personal interest had designedly overlooked."¹

¹ See also Simonde de Sismondi ([1826] 1991, 327–42). This became an important theme in the classical critique. "In the infancy of any such employment, it is only by actual wealth, in the shape of additional capital, that any effectual assistance can be given to a new branch of industry," Jeremy Bentham (1843 [1821], 96) stated. "By removal of competition, increase

Others were skeptical of the infant industry argument on the grounds that protection would just create inefficient industries. J. R. McCulloch thought it unlikely that the domestic industry could ever succeed in reducing its price below the lowest import price, but according to D. P. O'Brien (1970, 221) was apparently inconsistent on this point. George Scrope (1833, 369) believed that import prohibitions could "raise some faint imitation of a foreign manufacture in a country unsuited for its production," but that "there is a waste of all the trouble and expense which the effort has cost. . . . In the sickly and confined atmosphere of the legislative forcing-frame [industry] loses its health and vigour, decays, and before long expires." Jean-Baptiste Say (1834, 131) accepted Smith's dictum in general, but then conceded that cases could exist wherein "a new channel of industry may ruin an unsupported speculator, though capable of yielding enormous profit, when the labourers shall have acquired practice, and the novelty has once been overcome."

The debate over infant industries from the mercantilists through the classicals (but prior to Mill's statement) hinged on three key issues: whether infant industry protection would (1) create new wealth and capital, or merely divert it from other more profitable activities; (2) stimulate domestic producers to acquire new technology and skills, or just stifled the incentive for such efforts; and (3) generate long-term net benefits, or simply foster costly industries that required ongoing government support. On the first issue, the conceptual framework of Smith and Bentham was structured in a way that emphasized resource constraints, whereas proponents (coming from a different perspective) disputed or ignored that point. On the second issue, the stimulating effect of infant industry protection envisioned by Steuart ("promotes dexterity and emulation in invention and improvement") was contrast with the depressing effect envisioned by Scrope ("loses its health and vigour, decays, . . ."). The third issue, the intertemporal cost-benefit analysis, was a crucial but neglected aspect of the economic case for infant industry protection.

Unfortunately, there was no agreement on how to determine which of these perspectives was most appropriate. Indeed, economic analysis alone was of little assistance in evaluating these issues: one could envision the successful maturation of the infant, and yet one could also entertain the possibility of protection breeding inefficiencies; *a priori*, neither outcome could be dismissed. Thus, the debate amounted to the exchange of claims and assertions about which effects were more likely. Indeed, for many de-

may indeed be given to the rate of profit, if profit be the result of the newly directed labour: but it is only by the employment of capital, which must necessarily be taken from other sources, that this result can be obtained; the prohibition of existing rival establishments will not create that capital."

cares, the problem afflicting the infant industry argument was the lack of a substantive economic structure behind it. Without such structure, economic theory could not provide a satisfactory resolution to the debate over whether government intervention was justified.

With some hedging, the classical economists broadly supported the view that infant industry policies were not advisable. They based this verdict on their observations about tariff policies in practice and on their view (following Adam Smith) that policies of “preference or restraint” could not improve upon the economic outcome that arose from a system of natural liberty. But if Smith and his followers believed that their criticisms were sufficient to undermine the infant industry argument, they were sorely mistaken. A decidedly different view of the issue was taken elsewhere in the world. Economic observers in America, Europe, and in other industrializing countries of the day saw no reason for importing manufactured goods from wealthy Britain when their own country seemed to have the skills and resources necessary to produce such goods at home. Far from suffering any inherent and immutable cost disadvantage, all their fledgling manufacturers apparently lacked was the accumulated experience and expertise required to produce such goods more efficiently. Three major figures—Alexander Hamilton, John Rae, and Friedrich List—gave renewed force to the infant industry argument even after Smith’s scathing treatment.

As the first United States Secretary of the Treasury, Hamilton penned the famous “Report on Manufactures” in 1791. Hamilton objected to Smith’s doctrine that, if left to itself, industry would automatically take the most profitable course of development.² Hamilton ([1791] 1966, 266–67) spoke of “the strong influence of habit and the spirit of imitation—the fear of want of success in untried enterprises—the intrinsic difficulties incident to first essays toward a competition with those who have previously attained to perfection in the business to be attempted—the bounties, premiums, and other artificial encouragements, with which foreign nations second the exertions of their own citizens in the branches, in which they are to be rivaled,” all as obstacles to the establishment of new industries. Overcoming these obstacles “may therefore require the incitement and patronage of government.” Although Hamilton recognized that import restrictions would increase domestic prices, “it is universally true, that the contrary is the ultimate effect with every successful manufacture. . . . Being free from the heavy charges, which attend the importation of foreign commodities, it can be afforded, and accordingly seldom or never fails to be sold cheaper, in process of time, than was the foreign article for which it is a substitute” (286).

² Hamilton drew quite extensively on the *Wealth of Nations*, even when he disagreed with it. See E. G. Bourne (1894).

Although Hamilton's discussion of infant industries was more detailed than any previous writer and his argument received widespread attention, there is little that is fundamentally new in his analysis. But his study of policy instruments is quite insightful for this period. Hamilton compared the effects of four policies to promote domestic manufactures: protective duties, prohibitions, export taxes on raw materials, and pecuniary bounties (subsidies). For three reasons, Hamilton's preferred method of intervention was a subsidy: first, bounties have a "more immediate tendency to stimulate and uphold new enterprises, increasing the chances of profit, and diminishing the risks of loss, in the first attempts"; second, "bounties have not like high protecting duties, a tendency to produce scarcity," that is, higher domestic prices; third, bounties also promote exports and thereby extend the size of the market for domestic producers (299). Alternatively, import duties raise revenue, but assist producers in the domestic market alone with no direct effect on exports. Recognizing the fiscal constraints to providing subsidies, Hamilton proposed the practical compromise of levying import duties with the resulting revenue being used to finance bounties on domestic production.

John Rae, a Scotsman who had migrated to Canada, provided a more acute analysis of infant industries. Disputing the notion of a harmony between the interests of an individual and the interests of society, Rae directly attacked Smith's claim that tariff protection would, by reducing national income, reduce capital accumulation.³ But Rae framed his discussion of infant industries more in terms of the advisability of government assistance to the transfer of superior technologies from other countries. Rae (1834, 364) believed that the "general practical conclusion" about the desirability of government intervention to facilitate technology transfer must be granted. But, he added, the case "resolves itself into particulars, and the investigations of the political economist, would seem to be confined to the

³ Although his response to Smith is not entirely clear, Rae (1834, 381–82) appears to imply that even if income were reduced slightly, the intensity of accumulation would increase and thereby bring about more capital. "It is said capital can only augment by accumulation, and, as the interference of the legislator takes something from individual revenue, it must also take from the power to accumulate, and, consequently, instead of augmenting, must tend to diminish the sum of the capitals of all the individuals in the society, that is the national capital or stock. . . . The answer to this objection is, that the proceedings of the legislator may increase the absolute capital and stock of the society, the provision, that is, for future wants, embodied in the stock of instruments possessed by it, though they may not increase, and may even a little diminish its relative capital, or the sum which would be brought out by measuring those instruments with one another. That is the amount of the absolute capital of the society, which is the proper measure of the wealth of the whole, and of each individual, and that whatever augments it not only directly, and of itself, advances national wealth, but ultimately, also, does so indirectly, through the stimulus given to the accumulation principle, and the addition thence arising to relative capital." For a more detailed discussion of Rae's critique of Smith, see Brewer (1991).

tracing out, from the principles of his science, rules determining when the passage of any art is practicable, and when the benefits derived from it will exceed, or fall short of the necessary expense in effecting the passage.”

Rae suggested that technology transfer policies should be employed if domestic manufactures could be expected to produce the goods at the same or a lower price than foreign manufactures, but without explaining when such an expectation was reasonable. Three distinct advantages would result. First, the infant industry would save the costs of transporting the imported goods and thereby increase internal trade, generate additional technological improvements at home, and “so increase the absolute capital of the society.” Second, the dislocation of domestic production caused by import supply disruptions, which causes a “great waste of resources,” would be avoided. Finally, and most important, greater domestic production of the previously imported good would “stimulate invention and diminish the propensity to servile imitation.” As he put it: “Every useful art is so connected with many, or with all others, that whatever renders its products more easily attainable, facilitates the operations of a whole circle of arts, and introduces change—the great agent in producing improvements—under the most favorable form” (365). Quoting Hamilton approvingly, Rae suggested that new arts, by “their very existence in any society, gives a powerful stimulus to the ingenuity of its members.”

Therefore, Rae was broadly supportive of infant industry policies:

The legislator effects his purpose by premiums for successful individual imitations of the foreign article; by general bounties on the home manufacture; or by duties on that imported from abroad. . . . it having been made sufficiently apparent that nothing prevents the branch of industry in question being established, but the difficulties attending new undertakings, the want of skilled labor, and a sufficiently accurate knowledge of the properties of the materials to be employed in the formation of the new instruments, it is then proper to proceed to direct and general encouragements by bounties or duties. In this way real capital, and healthy enterprise are directed to the art, the difficulties attending its introduction overcome in the shortest possible space, and the commodities yielded by it are produced at less outlay, and afforded at a less price than that, at which they were before imported. (368)

Rae sounded a note of caution in that the legislator “is never justifiable in attempt to transfer arts yielding utilities from foreign countries to his own, unless he has sufficient reason to conclude that they will ultimately lessen the cost of the commodities they produce. . . . When there are circumstances particularly unfavorable to the practice of the art, and no countervailing circumstances particularly favorable to it, the first introduction of it must always cost the society high, and the subsequent maintaining of it will in all probability be a burden on the common industry and stock.”

Unfortunately, he added, “examples of injudicious conduct of the legislator from inattention to this particular have been not unfrequent” (367–68).

Although Rae is perhaps the most careful early analyst of infant industry protection, Friedrich List was by far the most popular proponent of protection in newly industrializing countries. A German political activist, writer, and sometime academic, List’s book *The National System of Political Economy*, first published in German in 1841, attained the status within protectionist circles that the *Wealth of Nations* had achieved among free traders. List based his study largely on historical judgments rather than on economic analysis because he rejected classical theory, which he thought suggested that free trade was always beneficial. Instead, List argued that the appropriate commercial policy of a country depended on its particular stage of economic development.

List and the classical school agreed on many issues, such as, for example, the importance of freedom and stability in promoting investment and other forms of economic activity. But, in List’s mind, two basic points separated his doctrines from those of the classical school. First, he attacked the “cosmopolitical economy” of Adam Smith and his followers which, in his view, wholly ignored the separate and distinct economic interests of a particular country in a world rife with conflict, fraught with insecurity, and seething with national identity. List accused Smith and his free-trade followers of examining only what is best for the world overall under conditions that he believed presumed an as yet unattained degree of international cooperation. List’s harsh attacks on the “cosmopolitan school” of Smith frequently distorted the position of classical writers, who were not romantic cosmopolitans neglectful of the national interest, and even List admitted in his preface that his attacks were exaggerated for effect.

Second, List (1885, 133) stated that “*the power of producing wealth is . . . infinitely more important than wealth itself*,” and deemed this contrary to the approach of the classical school. The power of production, by which he apparently meant the ability to reproduce and augment certain factors of production such as capital and skilled labor, ensured not only “the possession and the increase of what has been gained, but also the replacement of what has been lost.” He accused the classical school of taking a static view that valued only current wealth (exchangeable value) to the exclusion of factors that could be used to produce wealth. His ideas about productive powers were not obviously at variance with the views of the classical economists, however, because they also emphasized the importance of allowing capital and skills to accumulate. List believed that more attention should be devoted to production, however, because “production renders consumption possible” (233).

When it came to commercial policy, List endorsed many aspects of mercantilist doctrine. List (1885, 144) believed that the interests of merchants

did not necessarily reflect the national interest in the development of a country's powers of production: "The foreign trade of a nation must not be estimated in the way in which individual merchants judge it, solely and only according to the theory of values (i.e., by regarding merely the gain at any particular moment of some material advantage); the nation is bound to keep steadily in view all these conditions on which its present and future existence, prosperity, and power depend." List repeatedly stressed the fundamental importance of manufacturing, the benefits of which were economic and noneconomic and included greater security and independence, greater division of labor with its impetus to developing skills and accumulating capital, and the like. Thus, the commodity composition of trade deserved watchful attention. "It may be stated as a principle," List (1854, 77) wrote, "that a nation is richer and more powerful, in proportion as it exports more manufactured products, imports more raw materials, and consumes more tropical commodities."

For these reasons, government support for infant industries was essential. "The fact that manufacturing industry transforms into productive capital, wealth, and national powers, explains mainly why protection exerts so powerful an influence upon the increase of national wealth." List (1885, 144–45) argued that future benefits from establishing domestic manufactures would more than compensate for what he fully acknowledged would be the short-run economic costs of protection:

The nation must sacrifice and give up a measure of material property in order to gain culture, skill, and powers of united production; it must sacrifice some present advantages in order to insure to itself future ones. It is true that protective duties at first increase the price of manufactured goods; but it is just as true . . . that in the course of time, by the nation being enabled to build up a completely developed manufacturing power of its own, those goods are produced more cheaply at home than the price at which they can be imported from foreign parts. If, therefore, a sacrifice of *value* is caused by protective duties, it is made good by the gain of a *power of production*, which not only secures to the nation an infinitely greater amount of material goods, but also industrial independence in case of war. . . . A nation capable of developing a manufacturing power, if it makes use of the system of protection, thus acts quite in the same spirit as that landed proprietor did who by the sacrifice of some material wealth allowed some of his children to learn a productive trade.

List (1885, 226–27) strenuously disputed Smith's negative remarks about infant industries:

He wrongly maintains that the revenues of the nation are dependent only on the sum of its material capital. His own work, on the contrary, contains a thousand proofs that these revenues are chiefly conditional on the sum of its

mental and bodily powers, and on the degree to which they are perfected, in social and political respects (especially by means of more perfect division of labour and confederation of the national productive powers), and that although measures of protection require sacrifices of material goods for a time, these sacrifices are made good a hundred-fold in powers, in the ability to acquire values of exchange, and are consequently merely reproductive outlay by the nation. . . . He has not considered the influence of manufactures on the internal and external commerce, on the civilisation and power of the nation, and on the maintenance of its independence, as well as on the capability arising from these of gaining material wealth.

However, not all countries were well suited for such policies. Indeed, List's case for protection was carefully circumscribed. Because manufactures only flourish in temperate climates, tropical countries must never attempt to acquire manufactures through artificial means.⁴ And regardless of the stage of economic development, all countries should have free trade in agricultural goods and raw materials.⁵ "The system of protection," List (1885, 188, 309) wrote, "can be justified solely and only for the purpose of the *industrial development* of the nation. . . . Measures of protection are justifiable only for the purpose of furthering and protecting the internal manufacturing power, and only in the case of nations which through an extensive and compact territory, large population, possession of natural resources, far advanced agriculture, a high degree of civilization and political development, are qualified to maintain an equal rank with the principal agricultural manufacturing commercial nations."

List drew upon his reading of history in proposing the best commercial policy for countries at different stages of economic development: "History teaches us how nations which have been endowed by Nature with all resources which are requisite for the attainment of the highest grade of wealth and power, may and must . . . modify their [commercial] systems according to the measure of their own progress: in the first stage, adopting free trade

⁴ "A country of the torrid zone would make a very fatal mistake, should it try to become a manufacturing country. Having received no invitation to that vocation from nature, it will progress more rapidly in riches and civilization if it continues to exchange its agricultural productions for the manufactured products of the temperate zone. It is true that tropical countries sink thus into dependence upon those of the temperate zone, but that dependence will not be without compensation, if competition arises among the nations of temperate climates in their manufacturing industry in their trade with the former. . . . This competition will not only ensure a full supply of manufactures at low prices, but will prevent any one nation from taking advantage by its superiority over the weaker nations of the torrid zone" [List (1854, 75–76)].

⁵ List (1885, 324, 187) wrote: "Free trade in agricultural products and raw materials is useful to all nations at all stages of their industrial development." He believed that production of "provisions and raw materials . . . needs no protection, and in which the restriction of commercial intercourse must be disadvantageous under all circumstances to both nations—to that which imposes, as well as to that which suffers from such restrictions."

with more advanced nations as a means of raising themselves from a state of barbarism, and of making advances in agriculture; in the second stage, promoting the growth of manufactures, fisheries, navigation, and foreign trade by means of commercial restrictions; and in the last stage, after reaching the highest degree of wealth and power, by gradually reverting to the principle of free trade and of unrestricted competition in the home as well as in foreign markets" (115).

Unlike Rae or Hamilton, List did not mention any preference for subsidies over tariffs as the way to promote industry. As for the height of protective tariffs, List observed that "it may in general be assumed that where any technical industry cannot be established by means of an original protection of forty to sixty per cent and cannot continue to maintain itself under a continued protection of twenty to thirty per cent the fundamental conditions of manufacturing power are lacking" (313). List was prepared to wait decades before allowing protection to expire, calling it "ridiculous to allow a nation merely a *few years* for the task of bringing to perfection one great branch of national industry" (319).

List never disparaged the ultimate goal of worldwide free trade; indeed, he embraced it.⁶ He simply advocated temporary protective measures in countries passing through a certain stage of development to ensure that they could trade on an equal footing with more advanced countries in producing manufactured goods. As he (1885, 129, 131) put it, "The system of protection, inasmuch as it forms the only means of placing those nations which are far behind in civilisation on equal terms with the one predominating nation [England] (which, however, never received at the hands of Nature a perpetual right to a monopoly of manufacture, but which merely gained an advance over others in point of time), the system of protection regarded from this point of view appears to be the most efficient means of furthering the final union of nations, and hence also of promoting true freedom of trade. . . . In order to allow freedom of trade to operate naturally, the less advanced nations must first be raised by artificial means to that stage of cultivation to which the English nation has been artificially elevated."

Over the course of the nineteenth century, List came to have an immense popular impact. By the end of the century such eminent British economists as Alfred Marshall acknowledged and accepted many of List's basic ideas about infant industries in developing countries. But most economists did not accept the infant industry argument on List's terms and with reference to List's works. They were suspicious of his historical analysis, skeptical that it could provide a careful analysis of the problems faced by infant

⁶ List (1885, 122) simply believed that there were important political prerequisites: "If . . . we assume a universal union or confederation of all nations as the guarantee for an everlasting peace, the principle of international free trade seems to be perfectly justified."

industries or useful guidance about the conditions under which protection was or was not advisable. Under ideal circumstances, the implications of economic theory did not depend upon historical circumstances, as was the case with comparative advantage. And List did nothing to advance the theory underlying the infant industry case because this was not his purpose.

Thus, the infant industry doctrine did not gain formal acceptance into classical trade theory until 1848, when John Stuart Mill published the first edition of his *Principles of Political Economy*. The endorsement by an economic theorist of first rank was not so easily dismissed as similar statements coming from a Hamilton or a List. Mill's ([1848] 1909, 922) original statement was as follows:

The only case in which, on mere principles of political economy, protecting duties can be defensible, is when they are imposed temporarily (especially in a young and rising nation) in hopes of naturalizing a foreign industry, in itself perfectly suitable to the circumstances of the country. The superiority of one country over another in a branch of production often arises only from having begun it sooner. There may be no inherent advantage on one part, or disadvantage on the other, but only a present superiority of acquired skill and experience. A country which has this skill and experience yet to acquire, may in other respects be better adapted to the production than those which were earlier in the field: and besides, it is a just remark of Mr. Rae, that nothing has a greater tendency to promote improvements in any branch of production than its trial under a new set of conditions. But it cannot be expected that individuals should, at their own risk, or rather to their certain loss, introduce a new manufacture, and bear the burthen of carrying it on until the producers have been educated up to the level of those with whom the processes are traditional. A protecting duty, continued for a reasonable time, will sometimes be the least inconvenient mode in which the nation can tax itself for the support of such an experiment. But the protection should be confined to cases in which there is good ground of assurance that the industry which it fosters will after a time be able to dispense with it; nor should the domestic producers ever be allowed to expect that it will be continued to them beyond the time necessary for a fair trial of what they are capable of accomplishing.

Mill's standing and reputation among economists gave intellectual credibility to the infant industry argument for the first time. Economists and others who viewed free trade as the best policy for all countries regardless of the circumstances were dismayed by the respectability Mill lent to protection. Richard Cobden, the great free-trade activist in mid-nineteenth-century Britain, reportedly lamented on his deathbed that "I believe that the harm which Mill has done to the world by the passage in his book on *Political Economy* in which he favors the principle of protection in young communities has outweighed all the good which may have been caused by

his other writings.”⁷ And Alfred Marshall ([1890] 1925, 259) later remarked that “when John Stuart Mill ventured to tell the English people that some arguments for protection in new countries were scientifically valid, his friends spoke of it in anger—but more in sorrow than in anger—as his one sad departure from the sound principles of economic rectitude.”

Indeed, complaints soon reached Mill on how his statement was being distorted by protectionists to justify high tariffs in the United States, Canada, and Australia in the 1860s. In his correspondence, Mill condemned any general policy of protection—“an organized system of pillage of the many by the few,” he called it—but reiterated that, in principle, the infant industry claim to protection was valid.⁸ “It may sometimes be a good calculation for the future interests of the country to make a temporary sacrifice by granting a moderate protecting duty for a certain limited number of years, say ten, or at the very most twenty, during the latter part of which the duty should be on a gradually diminishing scale, and at the end of which it should expire.”⁹ But the complaints persisted and by the late 1860s Mill grew to question his own approval of the doctrine. To one correspondent Mill sighed, “But I confess that I almost despair of this general understanding [of the limits of the infant industry case] being ever practically established. I find that in Australia, protection is not advocated in this form or for this purpose, but that the vulgarest and most exploded fallacies are revived in its support.”¹⁰

Eventually, Mill recanted his view that import protection was an appropriate means of promoting infant industries, although he never abandoned his belief that such industries could exist and that this in principle constituted a genuine exception to free trade. “Though I still think that the introduction of a foreign industry is often worth a sacrifice, and that a temporary protecting duty, if it were sure to remain temporary, would probably be the best shape in which that sacrifice can be made, I am inclined to believe that it is safer to make it by an annual grant from the public treasury, which is not nearly so likely to be continued indefinitely, to prop up an industry which has not so thriven as to be able to dispense with it.”¹¹

Curiously, Mill never incorporated these views into later editions of the

⁷ Quoted in George Armitage-Smith (1898, 53).

⁸ *JSM*, XVII, 1798. References to Mill’s work are from *The Collected Works of John Stuart Mill* (Toronto: University of Toronto Press, 1965–91), hereafter referred to as *JSM* followed by volume and page number.

⁹ *JSM*, XVI, 1044.

¹⁰ *JSM*, XVI, 1420.

¹¹ *JSM*, XVI, 1516. “I am now much shaken in the opinion, which has so often been quoted for purposes which it did not warrant; and I am disposed to think that when it is advisable, as it may sometimes be, to subsidise a new industry in its commencement, this had better be done by a direct annual grant, which is far less likely to be continued after the conditions which alone justified it have ceased to exist” (*JSM*, XVI, 1520).

Principles. In the sixth edition of 1865, Mill (1909, 923) added the following passage to his infamous paragraph:

The expenses of production being always greatest at first, it may happen that the home production, though really the most advantageous, may not become so until after a certain duration of pecuniary loss, which it is not to be expected that private speculators should incur in order that their successors may be benefited by their ruin. I have therefore conceded that in a new country a temporary protecting duty may sometimes be economically defensible; on condition, however, that it be strictly limited in point of time, and provision be made that during the latter part of its existence it be on a gradually decreasing scale. Such temporary protection is of the same nature as a patent, and should be governed by similar conditions.

Then, in the seventh and final edition of 1871, Mill added the further qualification that replaced “a protecting duty . . . *will* sometimes be” with “*might* sometimes be,” and to add “*it is essential that* the protection should be confined. . . .”

In the decades after 1848, Mill’s qualified endorsement of infant industry protection failed to attract much support among economists. Mill’s paragraph succeeded in putting them on the defensive, but economists remained skeptical. Even Mill’s foremost disciple, John E. Cairnes (1874, 403), dismissed infant industry protection because “the inevitable result is that industry becomes unprogressive wherever it is highly protected.” Cairnes characterized Mill’s statement as the “*obiter dictum* of a great writer” and called attention to the “strict limitations” Mill had set down with his case. But he added that “with or without such limitations, however, I cannot but think that the position is untenable.” Henry Fawcett (1878, 111) agreed that Mill’s argument would be conclusive “if there were a reasonable probability that the conditions under which he supposes that such a protective duty could be imposed would ever be realized.” But it has been “incontestably shown” that “it is absolutely impossible to impose a protective duty under the stipulations on which Mr. Mill so emphatically insists.”

William Graham Sumner (1885, 117) believed that Mill’s case was “conceivable” and therefore not “absurd,” but maintained that “I strenuously dissent from Mill’s doctrine even as he limits it.” “Manufactures grow up as population increases and capital accumulates, and, in the natural order of industry, are best developed in countries of dense population and accumulated wealth,” observed Henry George (1886, 165) in rejecting the infant industry argument. “Seeing this connection, it is easy to mistake for cause what is really effect, and to imagine that manufacturing brings population and wealth.” Finally, J. S. Nicholson (1901, 364–65) commented that “temporary protection is impossible owing to the creation of

vested interests, that new countries especially require capital, which is repelled and diminished by protection, and that the artificial forcing of new industries is not advantageous in the long run." Echoing George, Nicholson argued that "the best way to promote the higher forms of industry is not by the simple process of exclusion, but rather by improvements in the education of the people."

Despite such skepticism, key questions about the infant industry case for protection remained unasked and unanswered. The first question related to identifying the specific market failures that give rise to the necessity for government intervention. The second question related to whether, given the existence of some market failure, the expected value of government action was positive. Yet these fundamental issues were virtually swept aside toward the end of the nineteenth century when economists were reacting against what they viewed as dogmatic *laissez-faire* arguments against government intervention that had become associated with political economy. A greater willingness to accept market failures and concede a greater economic role to the government served to insulate the infant industry argument from greater scrutiny, giving the infant industry doctrine the appearance of greater strength in economic analysis than it actually had.

In his widely read textbook, for example, Henry Sidgwick (1887, 489ff) of Cambridge University strongly endorsed Mill's passage and even made the incredibly bold claim that "the argument for temporary protection . . . is theoretically valid from what I have called a 'cosmopolitan' point of view," that is, from the standpoint of the world's overall economic welfare. In Sidgwick's words, "It is quite possible that the cost incurred may be compensated to the community by the ultimate economic gain accruing from the domestic production of the commodity previously imported; while yet the initial outlay, that would be required to establish the industry without protection, could not be expected to be ultimately remunerative to any private capitalists who undertook it." Here Sidgwick raised the first key issue: if the government recognizes that a currently unprofitable industry is potentially successful after a period of protection, what prevents private firms and entrepreneurs from borrowing in capital markets to cover their initial losses? Private action would fail, he argued, "if the difficulties of introducing the industry were of such a kind that, when once overcome by the original introducers, they would no longer exist for others or would exist in a much smaller degree." This implied some externality, but regrettably he did not elaborate.

Sidgwick instead offered the example of two regions equally suited for producing manufactures, one of which exported manufactures and the other agricultural goods. The region exporting agricultural products could then save on the transportation costs of imported manufactures by estab-

lishing domestic production itself; in other words, temporary protective duties would bring about a gain that would “consist chiefly” in saving the cost of transport.¹² Sidgwick did not explain why comparative advantage failed in this particular instance or how world welfare would increase as a result. This savings-of-transportation-costs argument for infant industry protection was not new (having been discussed by Rae and dating back to the mercantilists), and Sidgwick was essentially resurrecting the same case made by the well-known American protectionist, Henry Carey (1858).

Yet this savings-of-transportation-costs argument completely fails as a case for protection. From the 6th edition of his *Principles* in 1865, Mill (1909, 925ff) scornfully called Carey’s widely read argument for protection “totally invalid.”¹³ Mill strenuously rejected it and had, in fact, totally demolished it in the *Principles*. The Carey-Sidgwick doctrine completely ignored the insights of comparative advantage and failed to recognize that, in spite of the transportation costs, there is a savings to the country from procuring the good from abroad. As Mill (1909, 923) put it, “The burthen [of transportation costs] is borne for a more than equivalence advantage. If the commodity is bought in a foreign country with domestic produce in spite of the double cost of carriage, the fact proves that, heavy as that cost may be, the saving in cost of production outweighs it, and the collective labour of the country is on the whole better remunerated than if the article were produced at home.”

There had also been skirting around the issue of whether infant industry protection was justifiable on the basis of cost-benefit analysis. Charles F. Bastable (1887, 136–37), a professor at Trinity College (Dublin), conceded that the infant industry argument is “the most plausible case which can be made for protection.” But Bastable pointed out that government intervention could not be called a success simply because it enabled an industry to overcome a historical handicap and survive. Rather, a more decisive question was posed: “Will the certain and immediate loss, resulting from protection, be outweighed by the future gains from the new industry?” This question not only applied a basic cost-benefit test to infant industry policies, but, in Bastable’s opinion, the practical problems associated with such policies “strongly impress us with the belief that this special case is in reality no exception to the rule of freedom in international trade.” Bastable agreed that arguments for protection of manufactures on noneconomic grounds were prevalent, but on the demonstrable criteria of economic gain, something even Mill himself had been vague about, was more stringent.

Until Bastable’s statement, economists had noted the costs associated with infant industry protection, but failed to identify clearly the gains from

¹² As Jacques Melitz (1963) points out, transportation costs were an integral part of Sidgwick’s trade theory.

¹³ To one correspondent, Mill even wrote about Carey that his was “about the worst book on political economy I ever read” (*JSM*, XVII, 1589).

having an industry naturalized. If the rate of payments to factors of production was determined in other sectors of the economy and the infant industry merely drew these factors away from other activities, factors would earn no more than their opportunity cost and there would be no gain from having created a new industry. Rather, the industry that developed due to protection must create some quasi-rents (or producer surplus) for it to prove beneficial to the economy.

Once the costs and the benefits had been identified, then they could be compared. Paul-Gustave Fauveau (1873), a French mathematician and contributor to French economic journals, calculated the precise conditions under which infant industry protection would bring a net gain. Suppose a were the mean annual cost required to establish an infant industry and b were the mean annual benefits generated by the industry after infant industry protection was removed. If x were the number of years of infant industry protection, the present discounted value of the costs of the policy (assuming interest rate r):

$$a + \frac{a}{(1+r)} + \frac{a}{(1+r)^2} + \dots + \frac{a}{(1+r)^{x-1}},$$

which would sum to:

$$\frac{a}{r} \cdot (1+r) \cdot \left[1 - \frac{1}{(1+r)^x} \right].$$

The gains from the infant industry after year x would be an infinite sum of b that would amount in present value terms to:

$$\frac{b}{r} \cdot (1+r) \cdot \frac{1}{(1+r)^x}.$$

Equating the discounted stream of future benefits from the infant industry to the discounted sum of initial costs, such that there would be no net gain from the infant industry policy, yields the following relationship:

$$x \cdot \log(1+r) = \log\left(1 + \frac{a}{b}\right).$$

Assuming an interest rate of 5 percent, Fauveau calculated the mean annual “break-even” gain that would be necessary to compensate exactly for the loss from import duties. If protection lasts 5 years ($x = 5$), then the mean annual gains (in perpetuity) from the infant industry must be 28 percent of the mean annual loss to ensure no net gain or loss from the policy. If protection continues for 10 years, the mean annual gain must be 63 percent of the

mean annual loss; for 15 years, 108 percent of annual loss; for 20 years, 165 percent; for 30 years, 332 percent; for 50 years, 1,047 percent. This calculation demonstrates the cost-benefit hurdle that any infant industry policy has to surmount for the expected value of intervention to be positive. What economists including Mill had failed to do was to address the underlying economic structure of the infant industry argument in terms of specific market failures, specify the gains that accrue to a nation from naturalizing a certain industry, and then describe how these gains could compensate for the loss incurred while protecting the industry.

The infant industry argument received yet another lease on life, and again escaped careful scrutiny, when two leading economists of the next generation merely accepted the case in principle without dissecting it. Alfred Marshall (of Cambridge University) and Frank Taussig (of Harvard University) were both agnostic as to the underlying theory behind infant industries and pursued a case-study approach to assess such policies. In various writings, Marshall expressed his belief that the classical economists had been too dogmatically opposed to the infant industry argument and that protection in the newly industrializing countries of the United States and Australia was not an unmitigated evil.¹⁴ Marshall ([1903] 1926, 392) went so far as to say that “protection to immature industries is a very great national good” and that, while costly in some instances, “it would have been foolish for nations with immature industries to adopt” free trade. Marshall would often describe his visit to America in 1875 during which he concluded that the overall costs and benefits of protection were roughly balanced.

Taussig also fully accepted the infant industry argument when the impediments to establishing an industry were artificial, not natural or permanent. In studying the cotton textile and iron industries in the early nineteenth-century United States, Taussig (1883, 66–68) found that although “the conditions existed under which it is most likely that protection to a young industry may be advantageously applied . . . little, if anything, was gained by the costly protection which the United States maintained in the first part of this century.” As this constituted just one experiment, Taussig insisted that “the intrinsic soundness of the argument for protection to young industries is therefore not touched by the conclusions drawn from the history of its trial in the United States.” Indeed, he maintained that whether protection is beneficial or not is “simply a question of probability for the given case.”¹⁵

¹⁴ Phyllis Deane (1990) provides an overview of Marshall’s views on free trade.

¹⁵ Certainly there was the danger that, as Taussig (1905, 47) put it, protection, “so far from leading to improvements and eventual cheapening, leads to the retention of antiquated and inefficient methods of production.” But he also argued that ten years might be too brief a trial period for infant industry protection, with thirty years not necessarily being unreasonable.

Marshall and Taussig did nothing to advance or qualify the conceptual basis for the infant industry argument. While their own studies provided little support for infant industry policies in practice, their uncritical acceptance of the economic argument perpetuated the impression that the infant industry case constituted a valid and important exception to free trade. As A. C. Pigou (1906, 13), one of Marshall's prize students, declared, "Of the formal validity of List's [infant industry] argument there is no longer any dispute among economists," despite the fact that since Mill's statement there had really been no deeper examination of the precise theoretical issues involved.

Over the first half of the twentieth century, the infant industry argument remained a universally acknowledged theoretical exception to free trade, despite the continued skepticism among economists about such protection in practice. Over these fifty years, however, the theory behind the argument remained vaguely formulated. Not only was it difficult to lay down general rules for ascertaining the likelihood of infant industry protection being successful, it was not even clear what specific market failures and other underlying conditions gave rise to the infant industries in the first place. What was the specific barrier that necessitated government intervention: the lack of a domestic supply of skilled labor? insufficient accumulated production experience? failures in the capital market? Depending upon the particular factors that prevented infant industries from arising and maturing on their own, the implications for trade policy could be quite different.

When economists finally began to put more economic structure on the argument, the case for protection appeared more limited than previously thought. Because the argument was for temporary (rather than permanent) protection, the infant industry case was inherently dynamic and involved some irreversibilities, such as accumulated dynamic scale economies resulting from learning by doing. In this context, the focus turned to identifying the precise point of market failure because that determined which method of government intervention was best suited to improving upon a *laissez-faire* outcome. James Meade (1955, 255–57) described the infant industry argument as follows: if a firm tries to enter a market it will initially incur a loss, but "after a time experience would bring with it the necessary skills and know-how and the industry would turn out to be an economic one for the country to undertake." However, Meade pointed out, "infancy as such provides no argument even for temporary State support." If the firm can eventually earn a suitable rate of return, then private enterprise in the capital markets will have an incentive to provide these funds and "there will be no case for a State subsidy." Thus, if capital markets were efficient there would be no need for government action. And even if capital markets were not efficient, one would be hard pressed to explain why the government should not attempt to remedy the particular deficiency in the capital

market rather than impose trade restrictions (which would fail to solve the underlying problem).

An alternative market failure could lie in the acquisition of technical knowledge by the firm. Meade pointed out that “it may be difficult for one infant to learn without thereby affecting the knowledge of other infants.” In other words, an initial investment by one firm could affect the production conditions for all subsequent firms if that knowledge is easily transferable. If the knowledge generated by learning by doing or research and development expenditures spill over costlessly to other firms, then it is conceivable that no firm will undertake the initial investment. “In such a case the temporary subsidization of the first firm may be socially desirable; but this would be so not because infants have to learn but because infants teach each other.” The further qualification here is that spillovers must be geographically local or national in scope; if such knowledge spillovers were worldwide in nature, there would be no special barrier to domestic firms in acquiring that knowledge from firms in exporting countries.

The principle of policy targeting implicit in Meade’s analysis suggests that the most efficient government intervention addresses the failure of marginal conditions between price and cost to hold at the precise source of their divergence.¹⁶ In this vein, Robert Baldwin’s (1969) classic critique of the infant industry argument for protection stresses that, even if the precise market failure associated with infant industries has been specified, a trade policy intervention does not necessarily provide a remedy that will ensure the maturation of the infant. Baldwin noted that import protection alone fails to provide the right incentives for an infant firm to make additional investments in acquiring technological knowledge. Nor does it necessarily improve the firm’s ability to retain the benefits of its investments in knowledge, but does serve to make old production techniques profitable. The appropriate policy actions should focus on correcting the specific, underlying problems thought to be associated with acquiring new technologies and investments in new techniques.¹⁷ As Baldwin (1969, 303) justly concluded, “If the infant industry argument for tariff protection is worthy of its reputation as the major exception to the free trade case, it should be

¹⁶ This is the underlying logic of the theory of domestic divergences, discussed in chapter 10.

¹⁷ As Harry Johnson (1965, 28) explained, “Once knowledge of production technique is acquired, it can be applied by others than those who have assumed the cost of acquiring it; the social benefit at least potentially exceeds the private benefit of investment in learning industrial production techniques, and the social use of the results of such learning may even reduce the private reward for undertaking the investment. Where the social benefits of the learning process exceed the private benefits, the most appropriate governmental policy would be to subsidize the learning process itself, through such techniques as financing or sponsoring pilot enterprises on condition that the experience acquired and techniques developed be made available to all would-be producers.”

possible to present a clear analytical case, based upon well-known and generally accepted empirical relationships unique to infant industries, for the general desirability and effectiveness of protective duties in these industries.” Nearly a century and a half since Mill lent his support to the infant industry argument for protection, the case has still not achieved this level of intellectual coherence.

. . .

Despite the intuitively appealing metaphor of “infant” industries, analytical progress in assessing this argument for protection has come extremely slowly since Mill’s qualified endorsement. Although hampered by a lingering sense of vagueness, the infant industry argument in modern treatments now relates to hurdles faced by new firms in acquiring knowledge or capital. To the extent that government intervention is called for it is to improve upon the existing conditions of appropriability of knowledge investments or improving the functioning of capital markets. Trade interventions are not directly appropriate because those improvements may be desirable regardless of whether the industry in question is involved in international trade. As a result, this particular argument for protection is not nearly as prevalent or supportable today as it was several decades ago. Still, the infant industry argument has been difficult to dismiss altogether and it continues to occupy an uneasy place in the theory of commercial policy.

Chapter Nine

GRAHAM AND THE INCREASING RETURNS ARGUMENT

THE INFANT INDUSTRY argument for protection held that temporary protection might enable a domestic industry to reach a degree of efficiency such that it could, at some point, export without assistance at world prices. In the 1920s, Frank Graham, an economics professor at Princeton University, sought to describe conditions under which *permanent* protection could benefit a country. If manufacturing was subject to increasing returns to scale and agriculture was subject to decreasing returns to scale, then a country specializing in agriculture and importing manufactured goods could be depriving itself of production in a high productivity sector. Graham described the disadvantages of this situation and argued that a permanent tariff on manufactures could prove superior to free trade. As with previous arguments for protection, his case provoked sharp debate, but it foundered on the following criticism: increasing returns that are internal to the firm are incompatible with market competition. Still, Graham succeeded in bringing attention to the potential effects of external economies in generating arbitrary patterns of specialization and trade.

. . .

The classical theory of comparative advantage, which achieved prominence in chapter 7 of Ricardo's *Principles*, assumed one factor of production (labor) and constant costs of production: if the quantity of labor employed in the production of a good doubled, output would also double. Yet trade was often described in the context of several factors and different cost relationships. In 1815, Thomas Malthus, David Ricardo, and Edward West all independently applied the notion of diminishing returns to scale in the context of agricultural production: as more labor was applied to a fixed amount of land, the less each additional laborer was able to produce. Ricardo's famous critique of the Corn Laws skillfully used diminishing returns (or, alternatively, increasing costs of production) in agriculture to describe how rents, profits, and wages were determined and to assess the impact of trade policy on income distribution. Unlike Ricardo and Malthus, West also called attention to the possibility of increasing returns in manufacturing, in which the costs of production might decline as the scale of production increased. West (1815, 25) suggested that "the division of labor and application of machinery renders labor more and more produc-

tive in manufactures, in the progress of improvement.” This effect was also present in agriculture, but was more than counterbalanced by recourse to inferior land, a nonaugmentable factor of production, which gave rise to diminishing returns. West’s distinction between diminishing returns in agriculture and increasing returns in manufacturing was adopted by other classical economists and became established as a theme in economics that continues to the present day.¹ However, the implications of increasing returns for trade and commercial policy were not, at this point, drawn out.

Unlike the terms of trade and the infant industry arguments, the debate about increasing returns was the first in which the main participants were American economists rather than their British counterparts. Economists in the United States had long contended with or even advanced the casual argument made by protectionists there that import tariffs on manufactures were desirable to promote greater output in increasing returns industries. American economists also confronted the related belief that, when a domestic industry is characterized by increasing returns, an import tariff might not result in higher domestic prices. Rather, the popular argument reasoned, a tariff might actually reduce prices because domestic output could expand and serve the market at a lower cost. Francis Walker (1903) attempted with little success to illustrate that argument more formally. He used numerical examples loosely based on increasing returns and ostensibly designed to demonstrate the potential gain from tariffs if prices of traded commodities fell as a result.²

Thomas Carver (1902) did not reject the possibility of lower prices altogether, but at least recognized that if increasing returns at the firm level continued indefinitely the industry would be driven toward a one-firm monopoly. In this case, a tariff would simply strengthen this monopoly power and allow a higher price to be exacted from domestic consumers.³ But most economists were surprisingly lax about recognizing the potential difficulties that increasing returns might create for the theoretical case in favor of free trade. Frank Taussig (1927, 83), for example, agreed that increasing returns “may alter conditions under which international trade is carried on,” but nonchalantly dismissed them as “not of a novel kind, and therefore call for no new analysis.”

It took a British economist, J. S. Nicholson (1897, 308–9), a professor of political economy at the University of Edinburgh, to point out that increasing returns might create problems for the theory of comparative advantage.

¹ Nassau Senior (1826, 86), for example, proposed the following economic law, that “every increase in the number of manufacturing labourers is accompanied not merely by a corresponding, but by an increased productive power,” and asserted that if employment in the textile sector doubled, output would more than double.

² Jacob Viner’s (1937, 475) harsh but reliable judgment is that “his procedure is defective in almost every conceivable particular . . . [and] his results are totally devoid of significance.”

³ See also Alfred Marshall’s ([1890] 1925, 261–62) related discussion.

Nicholson envisioned the following scenario: suppose wheat production was subject to diminishing returns to scale (wherein the marginal product of labor declined as more labor was employed), while cloth production was subject to increasing returns to scale (wherein the average product of labor increased with employment). Specialization and trade would imply sharply different outcomes for two countries. The country having a comparative advantage in the production of cloth would enjoy both a rise in the average product of labor in cloth (as output increased) and in the marginal product of labor in wheat (as output contracted). This would lead to an increase in the wage rate; Nicholson, as others of his era were prone to do, implied that anything that increased labor's share in national income was inherently desirable.

The opposite occurs in the country having a comparative advantage in wheat: the marginal product of labor in wheat would fall as output expands, while the withdrawal of labor from cloth would decrease the average product of labor in that sector. If cloth sells at the same price per yard as wheat per bushel, then this process would usually stop when the labor used to produce 1,000 yards of cloth could be used to produce 1,000 bushels of wheat. However, with different returns to scale a country might use resources to produce less than 1,000 bushels of wheat because the standard analysis "takes no account of the reduction of the average yield to labour in cloth as the production is diminished." In this case, the extra labour in wheat could yield less than 1,000 bushels and the country "will obtain less cloth than before and the total national income will be less." For this country, which would completely specialize in wheat, Nicholson argued that the "only compensation would be in obtaining more cloth than before for the wheat exported," that is, a sufficiently large improvement in the terms of trade.

Nicholson's early but insightful analysis came just short of clinching his case because he neglected to illustrate directly the loss suffered by the country specializing in the diminishing returns sector; he only indirectly addressed the gains-from-trade issue and his argument about reduced national income was incomplete by itself. His awkward use of marginal analysis in agriculture and average analysis in manufactures signaled a potential problem in examining the international trade equilibrium with this sort of increasing returns. Marginal analysis in agriculture was standard: output would expand until a competitive equilibrium was established at which the (increasing) marginal cost of producing the good was equal to its price. This would not suffice in considering manufacturing: if marginal cost was a decreasing function of output, there would be no profits on the earlier, higher cost units of output when marginal cost was set equal to price. Thus, Nicholson was forced to couch his discussion of manufacturing in terms of average productivity (or costs): as output increased, the average cost of

production over *all* units of output would fall. This situation was potentially compatible with competitive equilibrium.

Curiously, Nicholson's example drew little contemporary comment or debate.⁴ In 1923, however, Frank Graham constructed a very similar example that appeared to demonstrate the loss suffered by the country that did not specialize in the increasing returns industry. Holding that comparative advantage is not an "infallible criterion" of the best commercial policy, Graham (1923, 200) argued that when manufacturing is subject to increasing returns to scale, "protection to manufactures may advantageously be continued much longer than would seem adequate to cover the infant stage, whether or not the industry could maintain itself without such aid." Proponents of temporary infant industry protection conceded too much, in Graham's view, because "it may be to a country's economic advantage to protect an industry which could not grow up or survive without protection and which never will be able to survive without it, an industry which has not comparative advantage when the protective duty is first levied nor ever attains one under it" (202–3).

To illustrate his point, Graham introduced a numerical example of comparative costs and trade in watches (subject to increasing returns) and wheat (subject to decreasing returns). Suppose in the absence of trade a country can use a unit of labor to produce 40 units of wheat or 30 watches. If the international ratio of exchange is 40 units of wheat for 35 watches, the country has comparative advantage in wheat: with trade and two units of labor, it could consume 40 units of wheat and 35 watches, 5 more watches than it could without trade. But the labor costs of production, under Graham's assumptions, are not invariant to the movement of labor between the two sectors. As labor shifts into wheat production and confronts diminishing returns, output becomes more costly to obtain and the marginal product of labor falls to 35 units of wheat; as labor leaves the increasing returns watch industry, output there also becomes more costly to obtain and labor can only produce 20 watches. If the international exchange ratio remains 40 units of wheat for 35 watches, the country at that particular point still benefits from trade. Without trade at that point the country could have 35 units of wheat and 20 watches; with trade it could have 30 units of wheat and 35 watches and, as 5 extra watches are worth more than 5 units of extra wheat, the country has a higher real income.

But that is an incorrect comparison, Graham noted. The correct comparison, he pointed out, is no trade (with labor producing 40 units of wheat or 30 watches) versus trade (with labor producing 35 units of wheat or 20

⁴ L. L. Price's (1898, 63) review of Nicholson's book merely noted that the author "shows the importance of the possible effects of the two laws (of returns) on the terms of exchange . . . his judgment on the theoretical foundation of protection is expressed with far more caution and discrimination than older advocates of free trade would have exhibited."

watches). With trade, the country gets 10 fewer units of wheat and 5 more watches than without trade, but this is a net loss because 10 units of wheat are worth more than 5 watches. Thus, Graham concluded, “At any given moment . . . it will pay [for the country] to specialize in wheat, but the final result of the specialization is to bring about a situation in which the citizens of [the country] get less reward for their efforts than if they had never carried on international trade at all” (207).

Graham recognized that if the costs he had assumed were marginal costs, then the harmful effects of specialization were not inevitable. In this case, the higher cost of producing wheat applies to the final unit of output, not to the previous (infra-marginal) units to the same degree.⁵ (In other words, the average product of labor is higher than the marginal product of labor, so simply multiplying the total amount of labor by its marginal product would understate total output.) But if the costs were average costs, then the probability of loss would depend upon the extent to which output expands and the rapidity with which costs rise as a result of trade. In the specific example considered by Graham, “the conclusion that [the] country under the conditions assumed must lose by free trade is inevitable.” This result provides the basis for “advantageous permanent protection”: because the country is better off in autarky or with less trade, it is “economically benefited by protection and may do well to keep it indefinitely” (208).

Although the country with comparative advantage in the decreasing returns sector would suffer losses, Graham also showed that free trade would increase overall world output. For the other country (or countries) specializing in the increasing returns industry, the amount of labor needed to produce both wheat and watches would fall with the opening of trade and they would reap an unambiguous gain. In producing watches, Graham repeated Nicholson’s use of average cost analysis. Once again, the cost decline could not be at the margin because that would imply no profits on the initial, higher cost units. Instead, the lower unit cost must pertain to the entire output; as Graham put it, “An extension of output will be reflected here in a lower unit cost for the whole product and not for the new increment merely, and all producers must have approximately equal costs or be put out of business by their competitors” (208–9).

Graham concluded that “it may well be disadvantageous for a nation to concentrate in production of commodities of increasing cost despite a comparative advantage in those lines; it will the more probably be disadvantageous to do so if the world demand for goods produced at decreasing cost is growing in volume more rapidly than that for goods produced at increasing cost” (210). He dismissed the idea that the terms of trade would im-

⁵ Of course, income distribution, in the form of higher rent to owners of land, would change in this case.

prove for the country specializing in the decreasing returns industry; that is, the price of wheat would rise relative to the price of watches because of the scarcity of resources used in the production of wheat and increasing returns in the watch industry. "There is no reason for supposing that this would happen," he maintained without explanation.

Implicit in these early analyses by Nicholson and Graham were two distinct problems with free trade under increasing returns. First, increasing returns implied multiple equilibria. Because the marginal product of labor was increasing with the amount of labor in the manufacturing sector, one could imagine one equilibrium in which no labor was devoted to manufacturing (and output in that sector would be zero) and another equilibrium in which all labor was devoted to manufacturing. Any allocation of labor in between these two positions could not result in a stable equilibrium because an additional unit of labor would contribute more in output than the previous unit of labor. A ranking of the desirability of these alternative equilibria remained outside of standard comparative-static exercise of showing the effect of protection on national income.

Second, although average cost analysis allowed a given intermediate allocation of labor to be a stable equilibrium, an economic inefficiency was also introduced. At any such equilibrium, average cost pricing implied that wages would be set equal to the average product of labor, not its marginal product as under perfect competition. In this case, too little labor would be employed in manufacturing (where the marginal product of labor exceeds the average product of labor) and output would be distorted from its optimum level, introducing a potential corrective role for protection.

But Graham's stunning contentions were swiftly challenged by Frank Knight, later a leading economist at the University of Chicago. Knight's (1924) counterattack to Graham's "fallacious" but "ingenious argument" rested on the compatibility of increasing returns and competitive analysis. Knight took issue with the following assumption that Graham (1923, 204n) had made in a footnote: "The reasoning in the text simply assumes that a decreasing unit cost is obtained by an expansion of the production of watches; whether the cause of it be external or internal economies is immaterial to the theory, tho it would, of course, affect the degree of its applicability."

The distinction between internal and external economies, casually dismissed by Graham, in fact proved critical to this debate. Internal and external economies, both forms of increasing returns, were introduced into economic analysis by Alfred Marshall in the late nineteenth century. Increasing returns are internal to the firm when the production costs of a particular firm decline as its own output expands. Increasing returns are external to the firm when the production costs of a particular firm decline as industry output expands, while its costs of production increase when its

own output expands. As Marshall (1920, 220–21) put it, internal economies (also known as economies of scale) are “dependent on the resources of individual houses of business engaged in [the industry], on their organization and the efficiency of their management.” External economies, by contrast, are “dependent on the general development of the industry,” some of which “depend on the aggregate volume of production of the kind in the neighborhood while others again, especially those connected with the growth of knowledge and the progress of the arts, depend chiefly on the aggregate volume of production in the whole civilized world.”

Knight’s first point was that increasing returns internal to the firm are not strictly compatible with competition in the industry.⁶ If a firm’s decline in costs was a continuous function of its own output, then one firm could serve the entire market at the lowest cost and the industry would become a monopoly. As he (1924, 597) put it, “If competition is effective, the size of the productive unit will tend to grow until *either* no further economies are obtainable, *or* there is only one establishment left and the industry is a monopoly. . . . When all establishments have been brought to the most efficient size, variation in total output is a matter of changing their *number*, in which no technical economies are involved.” Thus, Knight concluded, an “increase in the output of a commodity must increase its cost of production unless the industry is, or becomes, a monopoly.”

“The rejoinder to the above argument,” Knight continued, “is the doctrine of ‘external economies.’” But this “surely rests upon a misconception,” he added. “External economies in one business unit are internal economies in some other, within the industry,” Knight argued, a point to be discussed shortly. “Any branch or stage in the creation of a product which offers continuously a chance for technical economies with increase in the scale of operations,” Knight insisted, “must eventuate either in monopoly or in leaving the tendency behind and establishing the normal relation of increasing cost with increasing size.”

Knight also pointed out what proved to be a minor error in Graham’s two-country example. He correctly noted that, in the example given above, at least one of the countries would be completely specialized in one of the commodities, bringing to a halt any change in the marginal productivities of labor. If one country was still producing both goods, then the specialized country would trade at an exchange ratio equal to the cost ratio of the diversified country. In his reply, Graham (1925) conceded his mistake about the exchange and cost ratio and specialization and went on to demonstrate that it did not materially affect his results. But on Knight’s main contention, Graham missed the point and essentially restated that the distinction between internal and external increasing returns was unimportant.

⁶ This was recognized by Augustin Cournot in 1838 and by Alfred Marshall and others later, but was apparently overlooked by Graham.

In his own reply, Knight (1925, 331–33) reiterated what Graham had ignored: “What Professor Graham has to do to establish his theoretical position is to show that . . . industry really operates under decreasing cost, *without tending toward monopoly*.” Knight agreed that “if one can make his assumptions in regard to decreasing cost, [Graham’s] conclusion is correct.” However, “to vindicate decreasing costs in the sense required, it must be shown that there are, or may be, industries, in a condition of stable competition, in which no producer already engaged could decrease his real costs by expanding his output at the expense of other producers, and yet in which real costs would be decreased all around by new producers entering the industry in competition with those already there.” But new entrants had to bid factors of production away from other uses and perhaps had to employ factors of lower quality: “These inevitable sources of increasing cost must be more than offset by some kind of purely ‘external’ economies in organization,” Knight argued. “In spite of the weight of authority which may be cited for such economies, I have never succeeded in picturing them in my mind, or finding any convincing reason to believe they exist; and the hypothetical examples cited by Professor Graham in his reply has not assisted me in doing so.” Even if he could, “I cannot believe such conditions general enough to justify a special law in economic theory, or a special provision in tariff legislation.”

Despite these and other criticisms, Graham never recanted his case for protection. In a short book on tariffs published a decade later, Graham (1934, 81) continued to maintain (in a chapter entitled “Rational Protection”) that if a country did not specialize in an increasing returns industry it “might steadily be losing opportunities to improve its *per capita* general productivity and might even suffer an absolute decline therein.” For this reason, “protection for comparatively incompetent industries of declining cost per unit of output might then be warranted.” But by disputing the compatibility of (internal) increasing returns and market competition, Knight had undercut a big part of Graham’s case. As John Chipman (1965, 741) later wrote, “As long as Knight’s objection stood, Graham’s entire argument—whatever other defects it had, and there were several—was vitiated by having this as its premise.”

Taking their cue from Knight, other leading economists were equally skeptical of Graham’s argument. Like Knight, Gottfried Haberler (1936, 198–208) granted that “Graham’s conclusion follows, provided that one accepts his assumptions,” but he viewed those assumptions as “highly precarious.” Jacob Viner accused Graham of making the same mistake Nicholson had, that is, confounding marginal and average analysis. Viner (1937, 480) argued, “Had Graham dealt with his problem in terms of marginal costs and marginal returns for both industries, he could not have obtained results unfavorable to free trade.” This is because, Viner pointed out, and Graham had partially acknowledged, “specialization in accordance with

marginal cost to the industry of the country must be to the advantage of the country, in so far as costs are made the criterion of advantage” (474).

Jan Tinbergen (1945, 182–99) picked up this issue and noted another error in Graham’s analysis. Graham purported to show that the effect of trade is to obtain a smaller number of watches for the same effort as originally produced a larger number of watches via domestic production. But Tinbergen pointed out that the larger number of watches was calculated on the basis of the average (not the marginal) product of labor, whereas wheat production was computed on the basis of the marginal product of labor (which is lower than the average product). Graham had thus employed a bogus measure of foregone output and, as Tinbergen illustrated geometrically, if the correct marginal calculation had been made there was no loss from trade.

But Knight’s strictures had far less force if one accepted the concept of external economies. What precisely were these economies? As already noted, Alfred Marshall described external economies simply as cost reductions of an individual firm that are “dependent on the general development of the industry,” that is, on the scale of industry output. The relevant industry output could vary anywhere from local to national to world output; to repeat Marshall’s words, “some depend on the aggregate volume of production of the kind in the neighborhood while others again, especially those connected with the growth of knowledge and the progress of the arts, depend chiefly on the aggregate volume of production in the whole civilized world.”⁷ F. Y. Edgeworth (1905) first succeeded in rendering the concept of external economies compatible with competitive equilibrium, something that could not be done with internal economies. Under external economies, according to Edgeworth’s formulation, the marginal cost of an individual firm was increasing in its own output, but its cost curve would shift down with an increase in industry output. Because each firm was small relative to the industry, industry output was taken as given and the contribution of the firm’s own output to industry output was imperceptible. In this way, each firm alone would operate under increasing costs, but the industry supply curve might be downward sloped.⁸

⁷ For a discussion of Marshall’s ideas on increasing returns, see Renee Prendergast (1992).

⁸ An extensive debate during the 1920s and 1930s about the nature of these cost curves and the relevance of external economies (and diseconomies) takes us too far from the field of trade policy to merit review here. See the excellent discussion in John Chipman (1965, 736–49). In the end, a consensus was apparently reached that accepted external economies as compatible with competitive equilibrium. The social optimality of that competitive equilibrium, however, was disputed. Believing that the market-determined output in such industries would be too low, Alfred Marshall and A. C. Pigou argued that government subsidies were necessary to raise industry output to the social optimum. Pigou later recanted this recommendation in the case of internal economies after the attacks by Knight and Dennis Robertson, but did not disavow subsidies in the case of external economies.

As Knight's skeptical remarks indicated, the entire concept of external economies was never uncritically accepted, let alone its policy implications, partly because of Marshall's (and his followers') vague formulation of the idea. Dennis Robertson (1924, 24) expressed the confusion this way: an external economy really "only means that, given time and the progress of organisation, a larger output can be produced at a lower cost per unit than a smaller output used to be." But, he added, "we used not to dare conceive of falling cost per unit as a determinant of increased output, but only as resulting from it, or at the rashest as 'being associated with it.'" Robertson also noted that by mixing a discussion of static external economies with that of the dynamic elements of industry evolution and technological progress, Marshall's treatment was an invitation for further confusion. He questioned whether a subsidy or tariff was necessary to promote the development of external economies, inquiring, "is not the body (whether a private monopolist or State) which seeks to improve substantially on competitive output in such cases [of 'decreasing cost'] seeking to voyage *pennis non homini datis*, and not merely to penetrate the secrets of Time, but to do that leisurely old gentleman's work for him?"

The policy implications, of course, hinged specifically on what one had in mind as generating the external economies. As early as the 1870s, Marshall described three types of external economies: knowledge spillovers between firms, subsidiary supplier industries, and local pools of skilled labor. Knowledge spillovers, in which firms are thought not to appropriate fully the benefits of their investments in knowledge-generating activities, result in an unpaid side effect of one firm's activity on others and possibly the underprovision of that activity relative to the social optimum.⁹ Pigou interpreted this as an externality in which the marginal social benefit of knowledge production exceeds the marginal private benefit of this activity, calling for, in principle, a government subsidy to close this divergence (and thereby increase national income). However, knowledge spillovers are not necessarily related to static external economies because they need not be linked to the scale of current output.

The last two examples, by contrast, are more consistent with what are viewed today as external economies: namely, market-size effects, wherein

⁹ Marshall (1920, 225) described knowledge spillovers this way: "Inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes a source of further new ideas." He also implied that distance was becoming less important a factor in localizing knowledge. "For External economies are constantly growing in importance relative to Internal economies in all matters of Trade-knowledge: newspapers, and trade and technical publications of all kinds are perpetually scouting for him and bringing him much of the knowledge he wants—knowledge which a little while ago would have been beyond the reach of anyone who could not afford to have well-paid agents in many distance places" (237).

the larger the size of the market, the greater the productivity of the firms in the industry. For example, a large industry is better able to support more specialized suppliers of producer services, which in turn reduce the industry's costs and increase industry output. Marshall and his followers construed these to be external economies in the following sense: if a firm decides to enter an industry or increase its output, that firm does not take into account the fact that this entry and additional output will lower the cost of production for all other firms in the industry.

Yet Marshall's (1920, 264) description of this phenomena was characteristically vague: "The most important of these [external economies] result from the growth of correlated branches of industry which mutually assist one another, perhaps being concentrated in the same localities, but anyhow availing themselves of the modern facilities for communication offered by steam transport, by the telegraph and by the printing-press." Unfortunately, Marshall's followers were no more successful in coming up with compelling instances of such economies. Thus, it was never entirely clear whether such "obvious" instances of such economies ("transport developments, the telephone and the trade journal, the shop of the club and the market price, subsidiary industries, a skilled labour supply,—we have all at some time tried to memorise and to reproduce the formidable list," Robertson [1924, 26] remarked) were just descriptions of market interdependencies and evolution rather than true examples of market failures requiring some form of government action.

In some cases, such as transport developments, Knight's statement that external economies in one industry were internal economies to another rang true. In other cases, the external economy could be internalized through market mechanisms such as vertical integration or other pricing arrangements. But the whole concept of external economies still remained elusive. As E.A.G. Robinson (1931, 138) seemed to lament: "And so we chase this will-o'-the-wisp of external economy through industry after industry, and we find it vanishing in the end or absorbed in the economies of firms or organisations below their optimum capacity." And Robertson's quip, that policy action to exploit external economies would have the state merely trying to accelerate what time and progress of organization would naturally bring about, hardly made the case more compelling.

The elusive nature of external economies ensured that suspicions about the concept were never completely dispelled. Even more slippery were the implications for government policy, let alone free trade. Depending on one's interpretation, external economies could be so rare as to be essentially a *curiosa*, or could reflect a pervasive interdependence that is rife through every sector of the economy. This greatly confused the issue in terms of its ramifications for free trade. A particular external economy might indeed call for a government subsidy to encourage the underpro-

vided activity, and international trade might reinforce an undesirable direction of specialization. But the implications for commercial policy were still not obvious. Haberler (1936, 207), for instance, complained that external economies were so “vague and indeterminate in nature” and so “difficult to estimate [as to] their extent or value” that “it is not really practicable to base a policy of Protection” on their possible existence. The standard examples of external economies given in the 1920s and 1930s were described years later by Chipman (1965, 746) as being “so far-fetched that it is difficult to understand how an entire theory of commercial policy can be based upon them.”

Yet even if one accepted uncritically the possible existence of external economies, Graham’s case did not go unchallenged. Another critic pointed to the importance of reversibility of costs. Graham’s position was that protection had to be permanent because, if protection were to be removed, the increasing returns sector (being one of comparative disadvantage) would contract and costs would begin to rise again. Unlike the dynamic infant industry argument, in which the cost reductions are irreversible once the infant has grown up and protection has been removed, Graham’s case is static and therefore a permanent policy action is required to hold the economy in a particular equilibrium. Karl Anderson (1936, 167) believed that Graham’s argument was weakened if “the sources of decreasing costs when once discovered and utilized are not to be lost again.” Stating flatly that “nonreversibility is clearly the case with genuine external economies” because their presumed source, once in place, was not likely to be dislodged, Anderson concluded that “the basis for the kind of protection defended by Graham is all but completely destroyed.” Yet this conclusion is simply an assertion; durability remains a debatable issue. Even if it were true, he neglected to mention that Graham’s case then essentially reverts to the dynamic infant industry case.

But it was Viner (1937, 475–82) who delivered the key qualifications to the use of trade policies to take advantage of external economies. First, the implications for trade policy depended critically upon whether the external economies were a function of world output or domestic output. If the economies depended upon the size of the world market, then domestic firms would not be especially handicapped if some domestic firms reduced their output. For example, if the size of the world watch industry was affected by some special attributes of the world watch machinery industry, and such machinery were freely traded, there was no particular problem for domestic firms.¹⁰ If part of the domestic watch machinery industry contracts, all watch firms in the world are affected symmetrically by

¹⁰ In Viner’s (1937, 480) words, “If there is free trade in machinery, this economy in machinery costs will not be lost to the watch industry in a particular country merely because it is shrinking in size, if there is no shrinkage in the size of the watch industry as a whole.”

the lost economies, not just domestic ones specifically. This weakens the potential case for promoting domestic firms and directs attention to the tradeability of intermediate goods, a key point in modern treatments of external economies.

Second, Viner (1937, 480) took the position that, if the external economies were pecuniary, “then they are not real national economies and nothing is lost to the country when they disappear.” To clarify this comment, Viner’s (1931) distinction between “technological” and “pecuniary” external economies needs to be understood. Technological external economies described the interdependence of firms in a particular market in which the production function of individual firms was affected directly by the output of the industry as a whole; pecuniary external economies were those in which the profits of a firm were directly affected by the activities of other (upstream or downstream) producers. Tibor Scitovsky (1954, 145) later clarified these concepts and drew the implication that “technological external economies are the only external economies that can arise because of direct interdependence among producers and within the framework of general equilibrium theory.” However, Scitovsky spoke of the “scarcity of technological external economies” because “it is not easy to find examples from industry” of them. It is difficult to consider changes in the organization of production as technical externalities, for example.¹¹

Viner (1937, 480–81) concluded in this way:

A conceivable case for protection on the basis of the existence of external economies in an industry which from the individual producer’s point of view is at a comparative disadvantage in costs can be made out, therefore, only where these external economies are (a) dependent on the size of the national and not the world industry and (b) are technological rather than pecuniary, or, if pecuniary, are not at the expense of domestic sellers or services or materials to the industry. The scope for the application of the argument is extremely limited, especially as it seems difficult even to suggest plausible hypothetical cases of the existence of genuine technological external economies.

This reduced Graham’s example to, in his words, “little more than a theoretical curiosity.”

The decades after Alfred Marshall’s first exposition on external economies revealed that economists did not really understand how to think clearly about such effects, let alone draw out the implications for trade policy. At the same time it did not seem worthwhile to abandon the concept entirely because it appeared keenly relevant to the seemingly arbitrary na-

¹¹ According to Scitovsky (1954, 136), “pecuniary external economies clearly have no place in equilibrium theory” because such “all-pervading” interdependence through the market mechanism cannot be held to lead both to an optimal allocation of resources and to a failure to lead to that optimum.

ture of comparative advantage in certain instances. An oft used example from R.C.O. Matthews (1949–50) asks, why is it that Germany specializes in cameras and Switzerland in watches? What is the role played by initial advantages in the past that cumulate to the present? Do scale economies or external economies account for this concentration and localization? If so, could not the location of the two industries be reversed and does this not illustrate the potential (as discussed earlier) for multiple equilibria?

Although the debate over external economies and comparative advantage was never satisfactorily resolved in the immediate aftermath of Graham's article, the study of the trade-related aspects of external economies nearly evaporated until the early 1980s. By this time, analysis suggested that Graham's case, taken on his own assumptions about external economies, was largely correct. Arvind Panagariya (1981) described a small open economy with a decreasing returns and an increasing returns industry (via constant-returns-to-scale firms with economies external to the firm but internal to the industry, thereby consistent with perfect competition and average cost pricing). Under these assumptions, a permanent subsidy to expand output in the increasing returns industry (or conversely a tax to discourage output in the decreasing returns industry) is required to maximize national income. A country completely specialized in the decreasing returns industry may suffer a welfare loss after the opening to trade, although that depended upon whether the terms of trade moved in a favorable enough direction. As in the infant industry case, the optimal policy is an appropriately chosen subsidy and not a tariff which engenders the additional distortion to consumption choices. Related work by Wilfred Ethier (1982*b*) pointed out the implicit importance of country size in Graham's argument; the country specializing in the decreasing returns sector was less likely to lose by trade the smaller the country is and (paradoxically) the greater the degree of increasing returns.

But this work, and others like it, just assumed the existence of external economies. In the absence of more concrete evidence about such economies, or whether they arose inherently from a certain, plausible market structure, the practical importance of external economies for free trade remained open to question. Ethier (1982*a*) later provided a possible theoretical foundation for external economies in which each firm operates under constant returns to scale but the industry overall exhibits increasing returns. The example drew upon Marshall's suggestion that a larger industry can support a greater variety of specialized inputs to production at lower cost. Ethier showed the following: if there are economies of scale within each firm that produces a differentiated intermediate good (that is, monopolistic competition), then the aggregate production function of the industry that assembles these components exhibits the features of external economies.

The key issue again revolves around Viner's point about whether the intermediates are tradeable or not. If the intermediate goods are tradeable, then the external economy depends on international production of the final good, not domestic production, and there is no strong case for government action. If the intermediates are nontradeable, then the external economy is strictly national in scope and depends on the extent of domestic production. In the latter case, the size of the domestic market can determine comparative advantage; a larger market will support larger industries and hence have lower costs than those in smaller markets. Under this Ethier specification, as James Markusen (1990) has pointed out, external scale economies lead to several potential distortions, including multiple equilibria (with a high and a low level of output of the industry with external economies), average cost pricing, and inefficient factor combinations. A one-time production subsidy could be used to establish the right equilibria, and then a permanent factor subsidy could be used to correct for the inefficient factor combination. But the desirability of Graham's permanent protection is not so obvious.

. . .

Increasing returns in the form of external economies of scale still prove to be somewhat vexing: their effects seem potentially important and difficult to dismiss, yet our ignorance about them is still vast after a century of debate. With small differences in initial market size, trade could potentially reinforce a certain pattern of international specialization. But until the underlying factors giving rise to external economies are better understood, the case for a trade remedy of import protection as the obvious, first-best policy response has not been made. Our understanding of the determinants and effects of external economies is so weak, both conceptually and practically, that it has yet to be established whether they offer a reasonably clear case in which protection could enhance economic wealth. A greater division of labor may come about as a result of extending the market, but extending the market by artificial means does not necessarily generate a greater division of labor that results in scale advantages.