1   Introduction

Steven Brakman and Ben J. Heijdra

1.1   Introduction

In speaking of theories of monopolistic or imperfect competition as 'revolutions,' I know in advance that I shall provoke dissent. There are minds that by temperament will define away every proposed revolution. For them it is enough to point out that Keynes in 1936 had some partial anticipator in 1836. Newton is just a guy getting too much credit for the accretion of knowledge that covered centuries. A mountain is just a high hill; a hill merely a bulging plain. Such people remind me of the grammar-school teacher we all had, who would never give 100 to a paper on the ground that 'No one is perfect.' (Samuelson, 1967, p. 138)

Edward Hastings Chamberlin is the author of one of the most influential works of all time in economic theory – *The Theory of Monopolistic Competition*, which entered its eighth edition in 1962. Along with Lord Keynes's *General Theory*, it wrought one of the two veritable revolutions in economic theory in this century. (Dust cover text of Kuenne, 1967)

Although we stress the importance of the contribution by Avinash Dixit and Joseph Stiglitz (1977) throughout this book, the history of monopolistic competition is much longer than the past twenty-five years or so and goes back at least seventy years. The success of the Dixit–Stiglitz model of monopolistic competition might have come as a surprise to students of the history of economic thought, as it was by no means the first attempt to deal with imperfect markets or monopolistic competition. However, where the earlier attempts failed the Dixit–Stiglitz approach turned out to be very successful and has the potential 'for classic status' (see Neary,[1] chapter 8 in this volume).

In this introduction we will briefly review the two waves of literature on monopolistic competition theory, namely the one that started in 1933 and the one that commenced in 1977. The claim of this book is that the second attempt to model monopolistic competition was far more successful than the first, essentially because the second attempt introduced

---

We thank Avinash Dixit for comments on an earlier draft.

[1] According to Peter Neary, 'the first step on the road to classic status [is]: to be widely cited but never read. (The second step, to be widely quoted but never cited.)'

1

a formalisation that had all the relevant characteristics of monopolistic competition but was still relatively easy to handle.

This collection of papers will show that the re-formulation by Dixit and Stiglitz has contributed significantly to many areas of research; the main ones being international trade theory, macroeconomics, growth theory and economic geography. But even today the concept of monopolistic competition is not always appreciated. As David Kreps puts it in his influential micro textbook 'were it not for the presence of this theory in most lower level texts we would ignore it here altogether' (1990, p. 344). Kreps dismisses monopolistic competition as being too unrealistic, and challenges his readers to come up with at least one sector that could convincingly be described by monopolistic competition. This collection of essays, however, takes for granted that the Dixit–Stiglitz reformulation of monopolistic competition has become very successful, and asks why that is the case. This does not mean that the authors of the essays are uncritical about the model. The aim of this collection is to show why the model has become mainstream in such a short period of time and what we can expect from future developments regarding the modelling of imperfect markets.

This introductory chapter is organised as follows. In section 1.2 we briefly discuss the literature predating the first monopolistic competition revolution. This literature strongly hinted at the importance of increasing returns to scale and imperfect market forms but was unable to come up with a satisfactory model in which both phenomena could play a meaningful role.

In section 1.3 we briefly discuss (what we call) the first monopolistic competition revolution, namely the one that was started by Edward Hastings Chamberlin and Joan Robinson in the 1930s. We show that by the mid-1960s most (but not all) leading economists had come to the conclusion that the Chamberlin–Robinson revolution had essentially failed. In our view, there are two reasons for this lack of acceptance of the theory. First, the *timing* of the first revolution was unfortunate in that it coincided with the Great Depression and the emergence of the Keynesian revolution in macroeconomics. Second, and perhaps more importantly, Chamberlin and co-workers failed to come up with a canonical model embodying the key elements of the theory. It was not so much Chamberlin's ideas that were rejected but rather his *modelling approach* that was deemed to be unworkable.

In section 1.4 we turn to the second monopolistic competition revolution, namely the successful one that was started in the mid-1970s by Dixit, Stiglitz and Michael Spence. The timing of this second revolution was much better. The events in the world – the petroleum cartel, high

inflation, productivity slowdown, etc. – made the profession painfully aware of the limitations of the paradigm of perfect competition, and made it more receptive to theories that departed from that paradigm in all its dimensions, i.e. returns to scale, uncertainty and information, strategic behaviour, etc. In addition, the second revolution caught on because Dixit and Stiglitz managed to come up with a canonical model of monopolistic competition. We present a very simple version of the Dixit–Stiglitz model and show how it manages to capture the key Chamberlinian insights.

Finally, in section 1.5 we present a broad overview of the chapters in this book.

1.2   Precursory thoughts on imperfect competition[2]

By the end of the nineteenth century two market forms dominated the discussion of economic analysis, namely monopoly and perfect competition. The former assumes a single firm with exclusive control over its output and the market, resulting in profits that are larger than in any other market form. In contrast, the latter assumes a large number of sellers of a homogeneous product, where each individual firm has no control over its price. Free entry and exit of firms ensures that long-run profits are zero. Perfect competition was introduced to show that in some sense it is optimal and in fact represents an end-state, meaning that competition between buyers or sellers has come to an end and neither party can increase utility or profits by changing its behaviour. Changes occur only if exogenous variables change, but the question then becomes how fast and under what circumstances the new equilibrium will be reached. Competition might not actually lead to the blissful state but market forces are always pointing the economy in the right direction.[3] Monopoly by contrast maximises profits of the firm but from a social point of view is sub-optimal.

This state of affairs is reflected in Alfred Marshall's *Principles of Economics*, that presented these two market forms as the basic analytical tools to analyse markets. Other market forms are hybrids in between these two

---

[2] Our historical overview is rather succinct owing to space considerations. Interested readers are referred to Triffin (1940), Eaton and Lipsey (1989, pp. 761–6) and Archibald (1987, pp. 531–4) for more extensive surveys.

[3] As Arrow and Debreu showed, in general the conditions for a unique and stable (Walrasian) equilibrium are that (1) production is subject to constant or diminishing returns to scale, (2) commodities are substitutes (meaning that a price increase raises the demand for other products), (3) external effects are absent and (4) there is a complete forward market for all goods. Assumptions (1) and (3) in particular are dropped in monopolistic competition.

4   *Steven Brakman and Ben J. Heijdra*

polar cases.[4] Mainstream economics did not bother too much to analyse
imperfect market forms, because 'the large majority of cases that occur in
practice are nothing but mixtures and hybrids of these two' (Schumpeter,
1954, p. 975).

However, Marshall was aware that other market forms were not simple
combinations of perfect competition and monopoly. The special nature of
imperfect markets were conveyed to him in the form of the duopoly mod-
els developed by Cournot, Bertrand and Edgeworth in the second half
of the nineteenth century. The analysis of Cournot (1838) was particu-
larly important for him, as it handed him the apparatus to analyse market
forms in the first place. The problem with these models was that the re-
sults depended very much on special assumptions. Although Marshall did
not develop his own theory of imperfect competition, his awareness of the
so-called 'Special Markets' paved the way for later theories of imperfect
competition developed by Chamberlin and Robinson.

Notwithstanding some lip-service to the theory of imperfect compe-
tition, perfect competition dominated the analysis during this time and
other market forms were considered to be 'imperfect'. However, in perfect
competition, where each seller or buyer has no influence on market prices,
there is no longer room for individual competition, and forces leading to
industry growth are absent. The difficulty was then to reconcile the the-
ory of the market and that of the individual firm. Simple observation of
reality often contradicted the conclusions of (partial) supply and demand
analysis: diminishing returns for the individual firm is not an obstacle to
expand production. And average costs are diminishing at the point were
firms stop expanding output. This state of affairs troubled Marshall, as
decreasing (average) cost curves are incompatible with perfect competi-
tion. Marshall tried to solve this by introducing diminishing returns for
the individual firm (for individual firms, production factors are in fixed
supply), and external economies for the whole industry. The introduction
of external economies of scale at the industry level ensured that the com-
petitive equilibrium could be rescued. The central idea is that external
economies of scale create an interdependence between supply curves; the
combined supply of all firms reduces industry costs and ensures that the
combination of lower prices and increased supply can be an equilibrium.
External economies of scale are compatible with an industry equilibrium,
because an increase in demand will still increase the price for *individual*

---

[4] However, according to Schumpeter, Marshall 'had no theory of monopolistic compe-
tition. But he pointed toward it by considering a firm's Special Market' (Schumpeter,
1954, p. 840).

firms, as the marginal cost curve of each firm is upward sloping and each firm is operating at the minimum of its average cost curve. The price increase could stimulate new firms to enter the market, reducing (average) costs and raising combined supply. With internal economies of scale a market equilibrium is not possible as each individual firm can always undercut its rivals.

According to Marshall whether or not external economies could be encountered in practice depended on the general characteristics of an industry and the environment of the industry, like the localisation of an industry. In Marshall's words:

subsidiary trades grow up in the neighbourhood, supplying it with implements and materials, organizing its traffic, and in many ways conducing to the economy of its material . . . the economic use of expensive machinery can sometimes be attained in a very high degree in a district in which there is a large aggregate production of the same kind, . . . subsidiary industries devoting themselves each to one small branch of the process of production, and working it for a great many of their neighbours, are able to keep in constant use machinery of the most highly specialized character, and to make it pay its expenses. (Marshall, 1920, p. 225)

In modern jargon the linkages described in this quotation are so-called backward and forward linkages; the backward linkage is that firms use other firms' output as intermediate production factors, the forward linkage is that its own product is also used as an intermediate production factor by others.[5]

Furthermore, according to Marshall a thick labour market also benefits firms:

Employers are apt to resort to any place where they are likely to find a good choice of workers with the special skill which they require; while men seeking employment naturally go to places where there are many employers who need such skill as theirs and where therefore it is likely to find a good market. (Marshall, 1920, pp. 225–6)

These factors combined explain industry growth and show why:

the mysteries of the trade become no mysteries; but are as it were in the air . . . if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas. (Marshall, 1920, p. 225)

---

[5] The quote from Marshall merely seems to shift the problem to a different level, in the sense that external economies of scale in one industry must be explained by internal economies of scale in an upstream or downstream industry linked to it, and that raises doubts about sustainability of perfect competition in that other industry.

6   *Steven Brakman and Ben J. Heijdra*

For Marshall, however, his analysis of external economies created an additional problem, because he thought that internal economies of scale were at least as important as external economies (Blaug, 1997). In the presence of internal economies of scale the growth of an industry would benefit the largest firms (and create monopolies) and thus change the competitive forces within such an industry. Marshall had to introduce the concept of the representative firm to deal with this incompatibility. By introducing the representative firm, perfect competition and (external) economies of scale could be made consistent. But again in this case, as with perfect competition, strategic interaction between firms has been assumed away because firms are by assumption 'representative' for the whole industry.

But the consistency problems in Marshall's analysis of the market were not solved even by the representative firm. Marshall's famous period analysis assumed that in the long run the supply curve was a straight line. And this means that *in the long run* the volume of production of an individual firm is indeterminate: there is no unique intersection of the supply curve and a given price. So, Marshall's theory of perfect competition has no way of dealing with situations where the (long-run) marginal costs are constant (or declining in the presence of economies of scale). This state of affairs was most poignantly put forward by Sraffa (1926). According to Sraffa market imperfections due to returns to scale are not simple frictions, 'but are themselves active forces which produce permanent and even cumulative effects'. And he added yet another problem. Declining marginal costs would imply that the market is served by a single firm. But, according to Sraffa, in practice firms operate under declining marginal costs without monopolising the whole market. According to him, the combination of a declining supply curve and a negatively sloped demand curve limits the size of production. The idea behind a declining demand curve is that buyers are not indifferent between different suppliers. Each firm has his own *special market*; products are usually imperfect substitutes and have their own *special* characteristics.

In a sense Sraffa added to the confusion rather than solving the problem of combining increasing returns and the theory of market competition. The error Sraffa made was that he did not distinguish between price and marginal revenue, which was remarkable because the concept of marginal revenue had already been developed in a mathematical appendix in Marshall's *Principles*, in which he restates the monopoly theory developed by Cournot.[6] This was pointed out (again) by Harrod in

---

[6] Marshall casts his analysis in terms of net revenue, and only implicitly discusses marginal revenue. The concept of marginal revenue had to be re-invented (Robinson, 1933). This

1930.[7] For Marshall it was a minor issue and he did not make use of this instrument any further, because he did not need it in his analysis of perfect competition.

This was broadly speaking the state of affairs in the 1920s and 1930s. It was realised that the existence of economies of scale (of one sort or another) implied imperfect market forms, but it remained difficult to construct a satisfactory equilibrium concept for such imperfect market forms. On the one hand there was perfect competition, and on the other hand there was monopoly. Other market forms were considered to be some kind of hybrid of these two extreme forms of competition. So, one could suffice to analyse the two extreme cases in treating all other forms as an implicit mix of the two fundamental forms of competition. But no satisfactory theory of the market existed in which constant or declining marginal and average costs could be made consistent with market equilibrium. This led in the 1930s to a new theory of price determination. One can agree with Schumpeter (1954, p. 1150) that the confusion caused by Marshall was a very fertile one.[8] Marshall's analysis of the firm and economies of scale led him to develop the concept of the representative firm which invited a lively discussion on market equilibrium and returns to scale and this set the stage for the analysis of monopolistic competition.

1.3   Monopolistic competition in the 1930s

In 1933 two books appeared that changed the way economists dealt with imperfect competition, namely Joan Robinson's *The Economics of Imperfect Competition* and Edward Hastings Chamberlin's *The Theory of Monopolistic Competition*. Although Robinson revived the marginal revolution, in general Chamberlin is considered to be 'the true revolutionary' (Blaug,

is even more surprising considering that Cournot already used the concept of marginal revenue in 1838, and derived the familiar first-order condition for profit maximisation: marginal revenue equals marginal cost (Cournot, 1838).

[7] See Harrod (1967) for a review of his thoughts on this matter.

[8] Chamberlin, for example, attributed the origins and inspiration of his theory to the famous Taussig–Pigou controversy on railway rates which took place around 1900. This controversy was about the explanation of different railway rates. Taussig tried to fit different railway rates into the Marshallian theory of (competitive) joint supply by assuming that a unit rail supply is not homogeneous and that different demand elasticities for different stretches of railway result in different prices. In contrast, Pigou stated that it was not an issue of heterogeneity, but of monopoly coupled with the conditions necessary for price discrimination which could explain price differences. In general it is thought that Pigou won the debate.
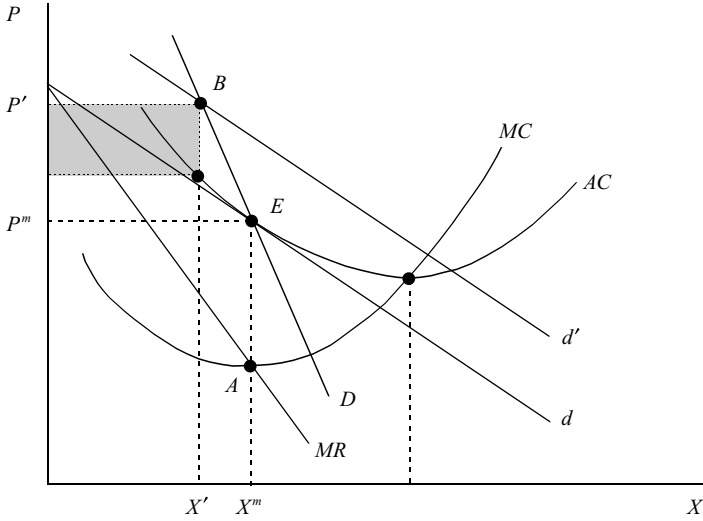
8   *Steven Brakman and Ben J. Heijdra*

Figure 1.1   Chamberlinian monopolistic competition equilibrium

1997, p. 376).[9] This radical new analysis was a first answer to the question that was raised in 1926 by Sraffa: is it possible in a market characterised by monopolistic competition and declining average and marginal costs to reach an equilibrium? Figure 1.1 illustrates the equilibrium in the monopolistic equilibrium. Chamberlin makes four basic assumptions (Bishop, 1967, p. 252):

- The number of sellers in a group of firms is sufficiently large so that each firm takes the behaviour of other firms in the group as given (Cournot–Nash assumption)
- The group is well defined and small relative to the economy
- Products are physically similar but economically differentiated: buyers have preferences for all types of products
- There is free entry and exit.

The monopolistic elements are all those elements that distinguish a product from another product and give the firm some market power; 'each "product" is rendered unique by the individuality of the establishment in which it is sold, including its location (as well as by trade marks, qualitative differences, etc); this is its monopolistic aspect' (Chamberlin, 1933, p. 63). The large number of firms in the market and the possibility of

---

[9] Moreover, the history of Chamberlin's seminal work dates back to 1921 – see the remarks by Schumpeter (1954, p. 1150).

entry and exit of many firms provides the competitive elements; 'Each [product] is subject to the competition of other "products" sold under different circumstances and at other locations; this is its competitive aspect' (1933, p. 63).

We illustrate the Chamberlinian model with the aid of figure 1.1.[10] We assume that all actual and potential suppliers in the group face the same demand and cost conditions and depict the situation for one particular firm in isolation. There are two demand curves in the diagram. The individual firm under consideration faces demand curve d. This curve represents the firm's price–sales combinations under the assumption that all other firms in the group keep their prices unchanged. Archibald calls this the 'perceived' demand curve (1987, p. 532). The steeper curve labelled D is the demand facing each firm if all firms in the group set their prices identically. Archibald (1987, p. 532) refers to this curve as the 'share-of-the-market' demand curve. As usual MR is marginal revenue (associated with the perceived demand curve d), AC is the firm's average cost, MC is marginal cost, P is the price of the differentiated commodity, and X is the volume of sales.

The Chamberlinian equilibrium under free entry/exit of firms is at point E, where the price is P^m and output is X^m. Point E is the equilibrium because (a) the individual firm attains an optimum in that point, and (b) there are no unexploited profit opportunities, excess profits are exactly zero and no entry/exit of firms takes place. The validity of these requirements can be demonstrated as follows. The individual firm maximises its profit, taking as given the demand curve d. It finds the optimum point by equating marginal revenue and marginal cost (see point A directly below point E). In point E the demand curve d is tangent to the average cost curve, AC, so the firm makes zero profits. This is the famous Chamberlinian tangency condition. Since all firms are identical, no firm makes profits or losses and there is no entry or exit of firms.

Chamberlin (1933, p. 91) also sketched the adjustment process towards the equilibrium point. Assume that all firms in the group are initially operating along the demand curve d' at point B, set a price of P', and produce a quantity X'. At this price–output combination, each firm would make a positive profit equal to the shaded area in figure 1.1. But point B cannot be an equilibrium. Indeed, in that point the individual firm will have an incentive to lower its price (and increase its profits) by moving to the right along the d' curve (recall that each firm operates under the assumption that its competitors will continue to charge P'). But each firm has exactly the same incentives, so they will all follow suit and cut their

---

[10] This diagram is adjusted from Bishop (1967, p. 252).

prices. As a result the d' curve will shift down along the D curve towards the Chamberlinian equilibrium at point E.[11]

Obviously, owing to the downward sloping individual demand curve, there is a difference between equilibrium average cost and minimum average cost in the Chamberlinian equilibrium. This implies that there are unexploited economies of scale and the question arises whether this represents a waste of resources. The answer to this question is both 'yes' and 'no'. 'Yes', in the sense that indeed there is excess capacity and 'no', in the sense that product differentiation introduces variety and this expands the extent of consumer choices and thereby welfare. As Eaton and Lipsey put it, 'in a society that values diversity, there is a trade-off between economizing on resources, by reducing the costs of producing existing products, and satisfying the desire for diversity, by increasing the number of products' (1989, p. 763). We will return to this topic in more detail when discussing the second monopolistic competition revolution.

Given the elegance of the monopolistic competition model it is surprising to see how little influence it had on economic theory. The first attacks on the early monopolistic competition revolution came from Hicks (1939, pp. 83–5) and somewhat later from Stigler (1949) and Friedman (1953). Hicks rejected the theory because he was unable to translate it into a workable model. Stigler (1949) rejected the theory for methodological reasons. He claimed that the predictions derived from the theory of monopolistic competition are not very different from those of perfect competition. Occam's razor then suggests that perfect competition should be favoured over monopolistic competition, a line of reasoning to which Friedman also adheres. It was put forward even more strongly by Archibald (1961, p. 14): 'The theory is not totally empty, but very nearly so' (see also Samuelson, 1967, for a further discussion of this debate). In addition Stigler raised an important point by noting that:

> Professor Chamberlin's failure to construct an analytical system capable of dealing informatively with his picture of reality is not hard to explain. The fundamental fact is that, although Chamberlin could throw off the shackles of Marshall's view of economic life, he could not throw off the shackles of Marshall's view of economic analysis. Marshall's technique was appropriate to the problem set to it: it deals informatively and with tolerable logic with the world of competitive industries and monopolies. But it is lost in the sea of diversity and unsystematism, and Chamberlin is lost with it. (Stigler, 1949, p. 22)

---

[11] Note that the position of the D curve depends on the number of firms in the group. In figure 1.1, D is consistent with the Chamberlinian equilibrium at E. As a result, the thought experiment conducted above does not prompt entry of firms. It just shows that E is the only conceivable Chamberlinian (Cournot–Nash) equilibrium.