« Barriers to Entry : Abolishing the Barriers to Understanding »,

Keppler, Jan Horst

Université Paris-Dauphine

2009

Jan Horst Keppler, « Barriers to Entry : Abolishing the Barriers to Understanding », *Journal of Public Finance and Public Choice* 27 (2-3), 2009, p. 99-124.

BARRIERS TO ENTRY: ABOLISHING THE BARRIERS TO UNDERSTANDING

by
Jan-Horst Keppler*
Professor of economics

Université Paris – Dauphine, LEDa, and
Université Paris I Panthéon-Sorbonne, PHARE

Port.: (+33 6) 77 81 37 46; Email: jan.keppler@dauphine.fr.

Abstract

The concept of a barrier to entry has been discussed least since Bain (1956) with important contributions by Spence (1977), Dixit (1980) or Milgrom and Roberts (1982). The more recent discussion is synthesized in the contributions to a dedicated session at the 2004 AEA meeting. Yet, a "barrier to entry" remains a surprisingly elusive concept, which even accomplished theorist fail to define in an unequivocal manner. This article shows that past and current contributions to the subject fail to resolve the issue, because they insist on analyzing barriers to entry in the context of identical firms in homogenous goods industries. This inevitably leads to logical inconsistencies as each theorist provides his own *ad hoc* definition of a barrier to avoid Bertrand competition. We show that the notion of a barrier to entry has economic pertinence only if interpreted as a fully sunk cost in the form of a unique, indivisible, non tradable factor of production. This however, implies monopolistic competition with welfare implications of a barrier to entry being positive as well as negative. Regulators thus need to ensure the ability of all competitors to create their own welfare-enhancing "barriers to entry" in a dynamic context rather than to focus on their existence in purely negative and static terms.

Key Words: Barriers to entry, sunk cost, irreversibility, monopolistic competition, regulation

JEL Classifications: B21, D21, D24, D43, G18

---

* University Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16 (Email: jan.keppler@dauphine.fr). I would like to thank participants at seminars at the University Paris-Dauphine and at La Sapienza University, Rome for their comments and encouragement. Particular thanks go to Domenico da Empoli, Richard Arena and Jérôme Lallement, all of whom have helped to sharpen particular aspects of the argument. Remaining errors of commission and omission are mine.

1. Introduction

The concept of a "barrier to entry" has been discussed at least since Bain (1956), with Stigler (1968) being a second major source for the large and growing literature on the subject. Among the many important contributions are the seminal articles by Spence (1977), Salop (1979), Dixit (1980), Milgrom and Roberts (1982) or Aghion and Bolton (1987). More recently, a fascinating session at the 2004 meeting of the American Economic Association posed the question "When are sunk costs barriers to entry? Entry barriers in economic and antitrust analysis" (see Carlton (2004), McAfee, Mialon and Williams (2004) and Schmalensee (2004)).

An attentive reading of this literature shows that the notion of "barrier to entry" remains a surprisingly elusive concept. The title of Carlton's contribution to the 2004 AEA symposium "Why Barriers to Entry are Barriers to Understanding" nicely recalls this fact. We will show that this elusiveness is due to the fact that the notion of "barrier to entry" as it is commonly employed contains both positive and normative aspects as well as static and dynamic aspects. Not separating these aspects clearly has lead to two major inconveniencies. First, it has lead to confusion and has held back the often inspired literature on the subject from converging towards a consistent approach to barriers to entry. Second, it has limited the usefulness of theoretical literature on the subject for practical policy in the fields of anti-trust and competition policy.

The present article will thus clarify and properly distinguish the different aspects that surround the concept of a barrier to entry. It will set out to do so in Section 2 by highlighting the fundamental difficulties of coming to terms with the issue of barriers to entry that leads to confusion between positive and normative analysis, static and dynamic perspectives and the choice of monopolistic or perfect competition as the relevant paradigm. Section 3 will review the different historic definitions and approaches to the issue and by presenting a new definition of a barrier to entry that unifies the two historic definitions provided by Bain (1956) and Stigler (1968). The proposed definition is complete, analytic and operational. Section 4 will concentrate on the modern literature on the subject and explore the dynamic nature of barriers to entry as well as its intrinsic links to the theory of monopolistic competition. Section 5 will conclude.

2. Positive or normative analysis, statics or dynamics, monopolistic or perfect competition?

Concerning the distinction between normative and positive aspects, the literature on barriers to entry has a tendency to characterize barriers to entry as negative to the extent that they impede the normative ideal of perfect competition and free entry. Negative welfare consequences have even been included into a number of definitions for barriers to entry, thus dispensing the authors from attempting any analytical characterization.[1] The statement "barriers to entry are bad" is often the least common denominator of different contribution to

[1] See, for instance, the definition by von Weizsäcker: "A barrier to entry is a cost of producing... that implies a distortion in the allocation of resources from the social point of view (see McAfee et al. (2004), p. 461-2 for additional examples)."

the subject. Characterizing certain kinds of barriers to entry, such as scale economies, as "innocent" or "innocuous" barriers, however, does not help either (see e.g. Salop (1979), p. 335) as they continue the normative approach under a reversed sign.

We shall show that if barriers to entry are to have any analytical (positive) meaning they need to refer to the intrinsic differences between the production functions of two firms. From a positive perspective, a barrier to entry thus constitutes the competitive advantage of one firm (usually referred to as the "incumbent") over another firm (usually referred to as an "entrant").[2] From a methodological point of view the confusion between normative and positive analysis arises by analyzing a situation of imperfect or monopolistic competition inside a paradigm of perfect competition. Of course, welfare could be improved if the entrant had the same (lower cost) production function as the incumbent, however this fundamental asymmetry is part and parcel of the competitive process.

Barriers to entry are neither good nor bad, as it would be absurd from the point of view of welfare maximization to prefer in the name of symmetry that incumbents have costs as high as entrants. If one wants to talk about barriers to entry in an analytically coherent and practically relevant manner, then one needs to abandon the normative ideal of perfect competition and work in a framework of monopolistic competition. Of course, the most prescient writers on the subject always felt this and we will see how subsequent generations have put great ingeniousness in working their way around this issue. The price to pay is the lack of a coherent theoretical approach that is probably unique in the science of economics for such a fundamental issue.

Closely related to the distinction between a normative and a positive approach is the distinction between static and dynamic analysis. Most of the literature on barriers to entry is an uneasy compromise between the two. On the one hand, virtually all authors on the subject agree that dynamic considerations are important (witness the "incumbent – entrant" dichotomy). However, most authors limit the dynamic aspects to developing multi-period games, in which only the strategic variables price and quantity can change (including the decision to enter a market or not). Barriers to entry, however, maintain a link to dynamic considerations in an even deeper sense. The existence of a barrier to entry implies firms being characterized by different production parameters, such as technology, product quality, vertical relationships *etc.*, which can change through historical time. We thus situate our own analysis in the line of argument reported by Salop:

> "In a personal communication Stiglitz pointed out that in general all deterrence instruments create intertemporal relationships in the profit function. In this sense, all deterrence instruments act as "capital," and a binding commitment corresponds to "irreversible investment (Salop (1979), p. 337)."

[2] In keeping with the literature on the subject, our analysis is limited to endogenous barriers to entry, i.e. barriers to entry over which a firm holds control, thus excluding exogenous barriers such as legal restrictions on entry. To some extent, the term "barrier to entry" is in itself problematic, since it suggests the existence of a barrier *independent* from other elements of a firm's production function. In addition, it maintains open the possibility that the industry in question is otherwise perfectly competitive. Of course, we have maintained the established nomenclature for reasons of readability. Nevertheless, we would like to point out that any notion of a barrier to entry other than an endogenously and consciously created factor of production by a participating firm is either not very interesting (such as the case of legal barriers) or will lead to contradictions (such as the distinction between "harmful" and "innocuous" barriers to entry).

Unfortunately, this important insight, which concludes Salop's 1979 paper, has never been elaborated on comprehensively in the subsequent literature. To a large extent, the present article is an attempt to fully develop its implication.

A second guiding thought for this paper is the insight by Ronald Coase that "…with zero transaction costs, the value of production would be maximized (Coase (1988), 158)." For our purposes, the following re-formulation is particularly relevant: "If all factors of production were tradable without transaction costs, then all sub-optimal allocations would be eliminated." As we shall see, the fact that certain factors of production are non-tradable is a necessary and sufficient condition for the existence of a barrier to entry. Alas, in that case the normative benchmark of perfect competition and free entry is no longer applicable. The reverse also holds, if all factors of production are fully tradable, in the sense that they are also divisible, new entrants can always buy capacity in optimal amounts and the outcome will be the competitive one.

Much of the existing literature on barriers to entry proceeds on the untenable double-assumption of non-tradability *and* assumption of free entry as a first-best optimum (see the discussion in Section 3). This leads to the almost habitual confusion between positive and normative aspects of the concept (see above). Making explicit the assumption that certain factors are not tradable, no longer allows naïvely confronting barriers to entry with the notions of free entry and perfect competition.

In keeping with the existing literature on the subject, we will only consider endogenous barriers to entry, *i.e.*, barriers to entry that are part of the strategic choice set of a firm actually or potentially participating in the market. Hence we exclude exogenous factors such as legal constraints. Most importantly, our approach will treat a barrier to entry as a *non replicable factor of production possessed by one firm but not by another*. This fundamental condition of non-replicability of the factor that constitutes a barrier to entry implies three essential qualities:

(a) The production factor in question is not tradable;
(b) The production factor in question is inseparable from the firm's overall production function;[3]
(c) The creation of the production factor in question is irreversible.

Although the three conditions imply each other, it is useful during the following discussion to be aware that all three of them apply.

From this basic notion of a barrier to entry as a non-replicable and non-tradable factor of production (and we will show that no other concept will allow to develop a consistent approach to the issue) follow a number of key insights that we would like to develop in this article. Among them are:

1. Sunk costs (irreversible investments) are barriers to the entry of a rival only to the extent that they lower an incumbent's production cost or engender a specific attachment of consumers to its products through product differentiation.
2. There exists no meaningful notion of barrier to entry other than firm-specific sunk cost with the exception, of course, external barriers to entry erected by third parties that are not considered in this article.[4]

[3] Thus a barrier to entry is not an external add-on to the production functions of otherwise identical firms that would engage in welfare-maximizing Bertrand competition if it was not for the barrier. In such a scenario, the barrier would be tradable and could no longer impede entry.

3. From the notion of firm-specific sunk costs it follows that the notion of a barrier to entry belongs to the domain of monopolistic competition theory. In fact, much of the imprecision and confusion in the literature stem from the attempt to make a concept intrinsically linked to competition between differentiate firms and products relevant to competition between undifferentiated goods.

It is relatively easy to show that some of the most celebrated contributions to the literature on the subjects (we will concentrate our analysis in the following on Spence (1977), Dixit (1980), Milgrom and Roberts (1982) and Aghion and Bolton (1987) have not, or only imperfectly, elaborated on the necessity that "barriers to entry" be sunk to the extent that they are inalienable and firm-specific. This is not to say that these contributions have not advanced our understanding of the subject, in particular, concerning the game-theoretic interactions of rival firms. The above mentioned omission nevertheless limits their pertinence in dealing with real-world barriers to entry and in providing guidance to anti-trust authorities.

In addition, we will elaborate on two broader themes. First, the notion of barrier to entry is difficult to pin down in terms of the time horizon that applies to its analysis. Such a time horizon would need to be longer than the short run, in which only variable factors can adapt, and shorter than the long run, in which all factors of production can be adapted and are hence, by definition, replicable. This vexing quality of barriers to entry was an issue with which historically monopolistic competition theory has done battle with (not terribly successfully) and is an additional difficulty to properly analyze the concept.

The second theme starts with the observation that identifying barriers to entry with non-tradable, firm-specific assets implies doing away with much of the *ex ante* normative connotations that have surrounded the concept. From this change in conceptual approach, however, does not follow that barriers to entry have no longer any implications for antitrust policy. On the contrary, it allows a more consistent formulation of the relevant questions. At the heart of the problem are now (1) the replicability of the factor in question and hence the dynamic nature of the problem and (2) the degree of symmetry between entrants and incumbents in generating their own "barriers to entry" conveying competitive advantage. The objective can no longer be to "remove all barriers to entry". Instead the objective will be to strike a balance between static welfare losses due to the absence of Bertrand competition and the gains welfare gains resulting from the search for and development of barriers to entry. Of course, practical antitrust policy is accustomed to make such tradeoffs. What we argue here, however, is that the theoretical literature on the subject has not been as helpful as it could have been in assisting antitrust authorities in motivating and structuring such tradeoffs.

Antitrust policy needs to be forward- rather than backward-looking; its focus should be the potential entrant not the incumbent. The ability of new entrants to develop their own "barrier to entry" (i.e., inalienable competitive advantage) needs to be the primary policy objective

---

[4] We agree strongly with the statement by Jean Tirole that "The distinction between 'fixed costs' and 'sunk costs' is one of degree not of nature (Tirole (1989), 308)." However, we insist equally strongly on the fact that fixed costs can form barriers to entry only to the extent that they are sunk. The same care that is employed in characterizing the gradual transition from a fixed to a sunk cost will need to be employed in characterizing the amorphous and to some extent transient nature of barriers to entry. This temporal element and the frequently half-way nature is, of course, one of the key reasons for the difficulty of analyzing barriers to entry. However, not taking on this difficulty and pretending that "barriers to entry" can be satisfyingly and completely defined as a concept of comparative statics has lead to the doubts about the concept that persist until today (see the discussion at the 2004 AEA Conference).

rather than the abolition of existing barriers to entry that are barriers only to the extent that they already confer competitive advantage.[5] The homespun but nevertheless deeply relevant advice to "never consider sunk costs" regains all of its relevance in this context. Another direction of anti-trust policy suggested by our approach is the "commoditization" of barriers to entry, i.e., the reduction of transaction costs and the enhancement of the tradability of assets (such as long-term contracts, for instance) through requirements of standardization and transparency. Long-term contracts, for instance, would no longer act as barriers to entry if their existence and conditions were publicly known and thus could be bought out by more efficient entrants (see also our discussion of Aghion and Bolton (1987) in section 3).

3. Barriers to entry: Two historic definitions and one new one

The difficulties surrounding the notion of a "barrier to entry" are neatly put in focus by the lack of one widely accepted definition. For reasons discussed below – and partly already highlighted by other authors – the two historic definitions of barriers to entry proposed by Bain and Stigler and their implications both have severe limits. These two definitions are either presented as mutually exclusive (Carlton (2004), p. 467) or one is considered defining a subset of the other (McAfee *et al* (2004) p. 462). Other definitions of barriers to entry add little of substance to the two historical definitions except that some authors highlight the negative welfare implications of barriers to entry as their defining characteristic.[6]

In Bain's definition a barrier to entry is something by which an "established sellers can persistently raise their prices above a competitive level without attracting new firms to enter the industry (Bain (1956), p. 5)." Stigler defines it the following way:

> "A barrier to entry may be defined as a cost of producing (at some or every rate of output) which must be borne by a firm which seeks to enter an industry but is not borne by firms already in the industry (Stigler (1968), 67)."

In the course of this article, we will argue that both definitions lead up to a common underlying definition from two different directions. In this view, both Bain and Stigler capture important aspects of barriers to entry, neither of them fully drawing the logical consequences from their observations.

The new, unifying definition of a barrier to entry that we propose here and which we will motivate during the course of our argument is the following:

> "A barrier to entry is constituted when a firm possesses a unique, indivisible factor of production for which no market exists and from which it draws a competitive

---

[5] Interesting additional issues arise when incumbents develop products with public good characteristics (the most relevant example is, of course, the creation of networks or any other form of intermediation). Strictly speaking from an abstract point of view, antitrust authorities would then need to clearly identify the private and the public good characteristics of the product in question and hand over responsibility for public good regulation to sectoral regulators. On a purely theoretical level, the issue would thus not pose a problem for an analysis of barriers to entry. In practice, however, it is often precisely the separation between the private and the public good characteristics of a product that poses the problem.

[6] See McAfee *et al.* (2004, p. 461-3) for, a synopsis of eleven different definitions.

advantage over actual or potential rivals either by being able to produce at lower costs or by being able to offer more attractive products."

Crucial in this context is the idiosyncrasy or the specificity of the factor of production, *i.e.* the fact that it cannot be reproduced identically at reasonable cost by any other company. In other words, there exists no market for the factor of production constituting a barrier to entry.[7]

Clearly, this definition is not contradicting but clarifying and unifying the two historical definitions. The existence of a unique, indivisible factor of production is, of course, the underlying cause for Bain's increased rate of return and for Stigler's cost advantage. The crucial innovation in the definition proposed here is the qualification that in order to make a barrier to entry operative, the factor of production in question cannot be re-sold or replicated. It constitutes sunk cost following an irreversible investment in its purest form.[8]

The proposed definition furthermore highlights several crucial aspects of barriers to entry which allow imbuing the concept with new analytical meaning:

1. A barrier to entry is not an exogenous add-on to a company's production set that can be bought by other companies or sold on a secondary market, but is an intrinsic element of said production function.
2. The fixed factor of production constituting a barrier to entry is necessarily indivisible. Divisibility would imply that its constituent parts can be sold or replicated. This indivisibility of the factor in question implies some degree of increasing returns to scale.
3. The notion of "barrier to entry" becomes meaningless under production with constant returns to scale. The existence of a barrier to entry is *always* a sign of monopolistic competition in the sense that the firm advantaged by a barrier to entry has some degree of monopoly power. The notion of a "barrier to entry to an industry" is a contradiction in terms to the extent that industry means competitive industry. The very definition of an industry implies absence of barriers to entry. Logically, the modern literature on the subject exclusively analyzes barriers to entry in single-firm "industries", alas without further exploring the issue.
4. The specific (idiosyncratic) factor of production responsible for a barrier to entry might intervene on the cost side or on the demand side. In the latter case, it will be responsible for a differentiated non-replicable product. In both cases, it will allow for some degree of monopoly power enabling the firm in question to raise prices above

---

[7] The existence of a non-tradable, "fixed" factor will give rise to increasing returns to scale and some degree of monopoly power. The opposite relationship also holds. As early as 1925, Piero Sraffa had pointed out that tradability of factors will imply constant returns to scale and perfectly competitive pricing:

> "If one assumes that all factors of production are used by a great number of industries (and hence that they are also perfectly transferable from one to the other), their remuneration, from the point of view of each industry is fixed and cannot be considered, from this particular point of view as a rent (Sraffa (1925), p. 33)."

[8] Other authors, such as Tirole 1989, have commented on the link between the notion of sunk costs and the existence of barriers to entry without exploring the resulting notion of asset specificity. The very title of the session at the 2004 AEA meeting "When are Sunk Costs Barriers to Entry?" poses the link in form of a question to which the correct answer, of course, is "Always, to the extent that they are truly sunk." None of the three authors of the session, however, develops the notion of asset specificity.

marginal cost. This double nature of a barrier to entry was already noted by Stigler (1968, p. 70).[9]

5. Barriers to entry have an uneasy position between the short and the long-run, which is part of the difficulty to come to analytic terms with them. Barriers to entry do not belong to the short run, *i.e.* they do provide a degree of monopoly power beyond the changing of variable factors or of relative prices. However, the resulting monopoly power is best thought of in terms of quasi-rent. The growing substitutability of the barriers to entry of competing firms over time will erode any monopoly power. Given that such barriers to entry are complex, indivisible, non-codified factors, they invariably open themselves to considerations of dynamic change.[10] They do therefore not belong to a Sraffian long-run, in which all factors of production are perfectly divisible and hence perfect competition under constant returns prevails, but to medium-term timeframe with which economic theory at all times has had difficulties to come to terms with.

Joe Bain and barriers to entry: three early misunderstandings

Following these introductory remarks it is useful returning to the two historical definitions themselves. Curiously, despite the fact that his name is so closely related to the concept, Bain does not really provide a handy definition of the concept of barrier to entry, or of "the condition of entry" as he was wont to call it. The closest he comes to such a definition, is in the following passage:

> "…the condition of entry may be evaluated by the extent to which established sellers can persistently raise their prices above a competitive level without attracting new firms to enter the industry (Bain (1956), p. 5)."

This is usually taken as Bain's definition of a barrier to entry. However, it is obvious that this definition describes the symptom rather than the cause, the result of a barrier to entry rather than the conditions responsible for it, a fact also noted by McAfee *et al.* (2004, 462). This lack of rigor obviously obfuscates the issue further. Nevertheless, definitions of barriers to entry based on their consequences for profits and welfare rather than on analytic characteristics still abound in the literature (see, for instance the definitions by Ferguson, Fisher, Gilbert or Carlton and Perloff cited in McAfee *et al.*).

In the following, Bain tries to capture the concept further by defining the three key terms used in the above definition "entry", "competitive level of price" and "persistently". Notably his definition of "persistently" allows developing an intuition for his idea of a barrier to entry:

---

[9] Of course, one can always argue that consumer preferences for the product of one firm over that of another can always be neutralized by a price differential. This takes us back to the question of cost. Nevertheless, the explicit introduction of the demand side is useful to highlight the potentially welfare-enhancing attributes of a "barrier to entry".

[10] One can think of barriers to entry as non-replicable informational assets such as, for instance, behavioral reputation ("aggressiveness"), brand names, patents, licenses, enterprise culture, management acumen, design prowess, pricing information or client relationships. In fact, practically all "fixed" factors responsible for the creation of increasing returns to scale are rooted in such informational indivisibilities (see Keppler (1998)). An important strand of the literature on the subject (see, for instance, the pioneering article by Milgrom and Roberts (1982)) thus concentrates rightly on reputation. The limits of this literature are that they concentrate exclusively on inter-firm reputation or signaling on pricing strategies rather than on the more important aspect on qualitative signaling vis-à-vis customers.

> "A third term that must be considered is "persistently"… The condition is inserted deliberately to give a long-term and structural aspect to our definition of the condition of entry… By a persistent elevation of price relative to a competitive level, we mean one maintained on the average over a substantial period of time, long enough to encompass a typical range of varying conditions of demand, factor prices and the like. Such a period might normally be thought of as five or ten years (Bain (1956), p. 7)."[11]

Bain views barriers as long-run structural features. An earlier passage already stated that "… the condition of entry is a structural concept (Bain (1956), p. 4)." However, by insisting to view a barrier to entry as a long-term structural feature Bain initiates a triple misunderstanding. This misunderstanding concerns (a) the intentionality of the company acquiring a barrier to entry, (b) the absence of dynamic considerations and (c) the domain of the impact of a barrier to entry.

First, as a structural feature of the competitive constellation, Bain poses a barrier of entry outside the intentionality of a firm's managers. This is plain wrong since there is nothing inevitable about barriers to entry. Every manager will strive as hard as he can to erect barriers to entry for his competitors. Bain's emphasis on "structure" is at the beginning of the notion of a barrier to entry as an *exogenously* imposed fact that will subsequently determine behavior and performance. Bain completely dismisses any feedback mechanism from performance to structure, as if a barrier of entry could not be *created* by the incumbent.[12] The circular definition of a "barrier to entry" provided by Bain and his followers as a feature causing higher-than-normal profits, negative welfare impacts) is the direct consequence of this structural approach.

Second, by defining a barrier to entry as a long-term structural issue, Bain *freezes* the concept in time. This detracts from the fact that barriers to entry due to unique, non-replicable factors of production are an essentially dynamic concept. Incumbents create and strengthen them, actual and potential incumbents are perennially in search for them. Competition between undifferentiated goods can only take place if the information sets of the competitors are identical. A barrier to entry instead implies monopolistic competition, where the differentiated information sets at the disposal of different competitors are in perennial evolution.[13]

---

[11] Bain's five-to ten-year timeframe might be interpreted in a way that a barrier to entry in a given industry has the typical lifetime of the length of an average capital-cycle. An alternative way of thinking about this issue might be a firm who possesses a patent on a valuable medical drug, whose profitability in excess of the normal return on capital will decrease once the patent elapses.

[12] The emphasis on "structure" can be easily employed to justify intervention by anti-trust authorities in order to remedy an imperfectly competitive situation that allows superior profitability for the incumbent. The issue becomes how to reduce existing "imperfections" (in the sense of a deviation from the normative ideal of perfect competition) rather than to encourage future "imperfections" that would allow superior performance through raising productivity, providing more attractive products etc.

[13] This is not to say that structural barriers to entry exogenous to a firm's decision parameters cannot exist. Stigler, for instance, cites the limited supply of New York City taxi licenses as an example of a barrier to entry (Stigler (1968), 67). This is the sort of external barrier to entry we choose to exclude from our analysis, even though the definition provided above would fully cover Stigler's example to the extent that a taxi license is indeed a non-replicable input to the provision of short-haul passenger transport services.

The third point concerns the choice of the term "barrier to entry" itself. By choosing this term rather than the more neutral term "competitive advantage", Bain puts the focus on firms that are not yet in the market, rather than on the incumbent. However, it is on the side of the incumbent that managers will strive to lower cost, attract customers and keep them loyal. Such product differentiation will erect the barriers to entry in question together with the concomitant scale economies. Even a "simple" cost advantage is the result of not yet generally available knowledge unique to the firm in question and hence a source of scale economies. We thus agree with Bain that scale economies are always part of a barrier to entry. What is missing in his reasoning is that these scale economies are due to the qualitative differentiation of a firm's productive process and/or its output from that of its competitors due to a non-replicable factor of production that is unique to the incumbent.

Bain, however, provides at least some indirect hints that barriers to entry might be the result of monopolistic competition. For instance in the "Preface" to his book he makes one direct reference to monopolistic competition theory:

> "The questions posed in such [this] inquiry have in the main not been the postulates of free-lance empiricists, but rather hypotheses stated or suggested by economic theory – particularly Chamberlinian and post-Chamberlinian price theory (Bain (1956), p. v)."

Unfortunately, this lead is never taken up again throughout the book and Bain never advances very far in a development of an analytic notion of barriers to entry that would go beyond the identification of a never fully specified, slightly mysterious, structural factor responsible for supra-normal profits.

One should, however, be careful not to judge Bain's work too harshly by looking at it through a prism of logical consistency and theoretical depth. Bain is an empiricist to the core and *was* interested more in the measurable consequences of "barriers to entry" (whatever they would turn out to be) rather than their causes. Establishing stable conceptual structures was not a primary objective of his work. His bibliography mentions besides Chamberlin's *Theory of Monopolistic Competition* (1933) only two works that can be qualified as theoretical, George Stigler's work on *Monopoly and Oligopoly by Merger* (1930) and E.A.G. Robinson's *Structure of Competitive Industry* (1932).[14] The rest are descriptive and statistical works on specific industries. What is surprising thus is not so much that Joe Bain did not deliver what he had never really aimed for but rather that the theoretical literature has not advanced more radically since then.

George Stigler and barriers to entry: an institutional impediment to say the right thing

George Stigler must be credited for the greatest part of the limited advance in coming to terms with barriers to entry since Bain's writings. His definition is more analytic than Bain's musings and allows identifying a barrier to entry directly rather than only indirectly by looking at its consequences for profits and welfare. The full definition provided by Stigler in Chapter 6, "Barriers to Entry, Economies of Scale, and Firm Size", of his *Organization of Industry* is:

[14] The latter work is particularly revealing in this context, as it constitutes a last-ditch attempt to defend the Marshallian competitive industry at a time when Robinson's Cambridge peers, most important among them his own wife Joan were already actively dismantling the concept in order to introduce their own version of monopolistic competition.

> "A barrier to entry may be defined as a cost of producing (at some or every rate of output) which must be borne by a firm which seeks to enter an industry but is not borne by firms already in the industry (Stigler (1968), 67)."

Stigler's great merit is to *internalize* the notion of a barrier to entry into a firm's production function. It is fascinating from the point of view of this article, arguing that barriers to entry are a notion intrinsically linked to monopolistic competition with process and product differentiation, to see the extent to which Stigler prepares that very argument. This is all the more surprising given Stigler's strong, explicit and well-documented aversion to monopolistic competition theory in general and to his main protagonist, Edward H. Chamberlin, in particular.[15]

On the other hand, we have every reason to believe that this aversion is the reason that Stigler did not fully spell out the implications of his arguments. In fact, he stops short of calling "competition between differentiated competitors" what it is – monopolistic competition. Instead, Stigler limits himself to spelling out six "implications" of his above definition:

> "1. Entrepreneurs have different abilities and hence average costs may differ between companies at all levels of production.
> 2. Capital requirements are *not* a barrier to entry.
> 3. Mergers yield only temporary profits (i.e. no barrier to entry here).
> 4. A Barrier to entry exists if one company has a "product that is more attractive to buyers." A small shoe manufacturer adapts to new fashions more quickly, or a large manufacturer of engines has a superior network of repair facilities.
> 5. "Free" entry is the absence of cost differentials between incumbents and entrants.
> 6. Product differentiation is often treated as a barrier to entry. [FN] [The footnote reads: See J. S. Bain *Barriers to Entry* (Harvard University Press, 1956), chap. iv.] This is correct usage, on our approach, only if the costs of differentiation (design, advertising, etc.) are higher for a new firm than an existing firm (Stigler (1968), 70)."

All six implications of Stigler's definition of a barrier to entry imply monopolistic competition with various degree of precision. Let us discuss them one by one:

Implication 1: "Entrepreneurship" of varying ability is perhaps the best-known example in the literature on monopolistic competition of a unique, non-replicable factor of production for which no market exists and that affects a companies fixed cost (witness the discussions of Nicholas Kaldor, Edward Chamberlin and Ronald Coase on this subject in the late 1930s).

Implication 2: Pure capital requirements, in the absence of any differentiation of the investment that the new capital will finance, are indeed no barrier to entry. Sheer size is thus not sufficient to establish a barrier to entry. (See also below the discussion on the possibility of implicit collusion in non-differentiated industries.) We might add, however, that capital becomes a barrier to entry to the extent that it is "sunk", *i.e.,* the moment that no longer any market for it exists.

---

[15] See, for instance, Stigler's article "Monopolistic Competition in Retrospect" (1949) and its discussion in Keppler (1998). Here we might only recall Stigler's recollection of a personal meeting with Edward Chamberlin in his memoirs:

> ... when I was a professor at Columbia University, I attended a meeting of the American Economic Association in Washington, D.C., and on the flight back to New York to my surprise I found myself sitting next to Edward Chamberlin. He opened the conversation, 'You and Professor Knight are the two most mistaken economists I know on the subject of monopolistic competition.' Thank heaven it was a short trip (Stigler, 1988, p. 58).

Implication 3: This is a restatement of Implication 2 that sheer size is not sufficient to establish a barrier to entry.

Implication 4: A "product that is more attractive to buyers" is, of course, the original definition of product differentiation and hence the basis of monopolistic competition.

Implication 5: This is a restatement of the original definition in negative form. One might want to add that "free entry" not only implies the absence of cost differences between incumbents and entrants but also the absence of fixed costs, the divisibility of all factors of production and production under constant returns to scale.

Implication 6: This is the most mysterious passage of Stigler's chapter on barriers to entry. On the one hand it accepts product differentiation (and hence monopolistic competition) not only in substance but also in name, as a barrier to entry (although ascribing it to Joe Bain, who never mentioned beyond the reference to Chamberlin cited above). On the other hand, Stigler wants to reduce product differentiation to cost differences. This is not without problems. A new entrant might well spend *less* on advertising than the incumbent. To the extent, however, that his advertising and his product better capture consumers' hearts and minds there will be no barrier to entry. Stigler's argument would thus hold only under the assumption that a firm's unique asset (the attractiveness of its product to consumers in this case) can be perfectly replicated.[16] This implicit assumption allows him to move away from product differentiation and to save his definition for competition with undifferentiated goods. This subterfuge, however, produces only a virtual victory. In final instance, the distinction between the differentiation of production processes (by cost) and of products (by quality) is nugatory. In both cases, producers possess a unique, non-replicable asset that provides a degree of monopoly power and makes them monopolistic competitors.

Stigler's general approach is driven by the implicit postulate that differences in product quality and hence in consumer satisfaction can be translated into cost differences. Either the same standard of product quality can be achieved by spending more or the consumer can be compensated adequately through a price rebate. Separating thus between cost differences on the one hand and quality differences and hence product differentiation on the other, Stigler avoids that his definition straightforwardly embraces monopolistic competition as the underlying cause of barriers to entry. In the concluding paragraph of his chapter, he tries to justify this procedure with the words:

> "If the demand conditions of the firm vary with a characteristic associated with firm size, our language permits us to say that the characteristic is an element of demand or of economies of scale. In keeping with the general tradition of keeping out of demand as much of producer-controlled factors as one can, we would call these selling advantages economies of scale but no important principle is at stake (Stigler (1968), 70)."

Well, at stake is only whether, in keeping with general tradition, one refers to selling advantages as product differentiation or whether one tries to avoid the concept (or rather the word) by hiding it behind unspecified "economies of scale". Nevertheless, Stigler confirms here a crucial point made above: barriers to entry and monopolistic competition (even if one wants to avoid the term) make the distinction between producer-controlled factors (parameters of the production function) and the characteristics of demand for the products of a given firm no longer possible, as the two inexorably imply each other.

[16] In other words, a product's attractiveness to consumers would be a function solely of the amount of undifferentiated advertising expenditure.

Stigler's definition of a barrier to entry as a cost advantage has the historic merit to *internalize* the notion into a firm's production function. However, his refusal to include intrinsic characteristics of the resulting good explicitly (rather than implicitly and by example) in the definition was bound to create misunderstandings. It is not that the Stigler definition is wrong but it is incomplete, for reasons that having nothing to with the internal logic of his argument and everything to do with his well-known opposition to monopolistic competition theory. Product differentiation, barriers to entry and monopolistic competition also exist in the academic marketplace.

Formulating product differentiation in terms of the opportunity cost that an entrant would have to forego in order to create an identically attractive product and thus including it into Stigler's definition *ex post* would not solve the problem. In case the entrant does not succeed, product differentiation remains. In case he succeeds, two identical competitors will compete with identical goods and the barrier to entry has vanished. Once again, the notion of a barrier to entry requires the incumbent to possess an inalienable, non-replicable competitive advantage. Stigler's examples show that he clearly sensed this but was unable to bring himself to state it explicitly. The perplexity this procedure provokes is captured perfectly in the reaction by Denis Carlton:

> "Barriers to entry as defined by Stigler is clear, but perhaps strange, because the words mean something other than would naturally come to mind (Carlton (2004), p. 469).[17]"

In the end, despite his efforts to draw a line, George Sigler's definition and his subsequent explanations squarely place the notion of barrier to entry in the realm of monopolistic competition theory. For those familiar with his scathing dismissals of the subject this is itself an interesting fact. More importantly, however, Stigler develops the basis for a robust approach to problems of entry, even when only implicitly admitting differences in product quality and product differentiation.

4. Economies of scale, sunk costs and demand: the modern literature on barriers to entry

Having discussed the historic definitions of what constitutes a "barrier to entry", we would like to re-iterate once more the definition that logically flows from that discussion:

> "A barrier to entry exists when a firm possesses a unique, indivisible factor of production for which no market exists from which it draws a competitive advantage over actual or potential rivals either by being able to produce at lower costs or by being able to offer more attractive products."

Crucial in this context is the idiosyncrasy or the specificity of the factor of production, *i.e.* the fact that it cannot be reproduced identically at reasonable cost by any other company.

The "absence-of-market" condition implies that a notion of indivisibility is associated with a barrier of entry, as well as the absence of its duplicability. If the barrier of entry was a factor of production was divisible, its structure would be known it could be sold and replicated. It is hence best to think of a barrier of entry of an intrinsic, inalienable characteristic of the firm in

[17] « Clear, but perhaps strange » is in itself a curious expression. And what is different from « what naturally come[s] to mind » is not so much the meaning of the words Stigler uses than rather the intention with which he uses them.

question, its DNA, to use a modern analogy, which makes it unique. Other characteristics associated with the "absence-of-market" condition are irreversibility of investment and the presence of sunk costs. The fact that costs are sunk is identical to the notion of "commitment" that comes with an irreversible, non-tradable investment in capacity, which is central to the important Stackelberg-Spence-Dixit model (see below).

If barriers to entry are due to an indivisibility in the production function, then production under excess capacity will be the norm. The existence of a unique production function including an indivisibility implies at least some degree of product differentiation and monopoly power. In other words, the notion of a barrier to entry is anchored intrinsically and inevitably in the domain of monopolistic competition. "Factor differentiation" in form of an unique factor of production has economic relevance only to the extent that it reduces either the cost or enhances the utility-relevant characteristics of the product. *There is no such thing as a barrier of entry between identical firms producing identical goods*. Not observing this simple insight is the source for most of the "barriers to understanding" mentioned by Carlton and the resulting confusion surrounding the concept.

What is more, given that the fixed factor in question is not replicable and hence its products are no longer identical to those of a rival, firms will produce under down-sloping demand curves with excess capacity and increasing returns to scale.[18] The fact that excess capacity is an intrinsic feature of barriers to entry was first explicitly stated by Spence (1977). However, instead of establishing itself as the relevant paradigm for the analysis of barriers to entry, the excess capacity approach was quickly shelved following the critique of Dixit (1980). See the discussion below.

The importance of increasing returns to scale also vindicates, of course, Joe Bain's insistence on scale economies (see above). There are two important qualifications to this. First, considering such scale economies "structural" facts rather than resulting from firm-specific innovation ties the concept again to a notion of static equilibrium, which inevitably leads to the circular reasoning so prevalent in the literature on the subject that "barriers to entry are present when profits are higher than average".

Second, in opposition to Bain one needs to insist that indivisibilities, scale economies and barriers to entry *can* but *do not need* to lead to excess profits. Excess profits depend on the specific relations of substitutability between differentiated goods facing a demand curve that is not fixed but may expand if a greater variety of products is offered. Thus, the "absence-of-a-market" for a specific factor is a necessary but not a sufficient condition for a barrier to entry. While a barrier to entry will always be due to factor specificity, factor specificity alone will create a barrier to entry only to the extent that it allows producing cheaper and/or more attractive products. The existence of a barrier to entry in the sense of limiting a competitor's access to a market relies on the existence of sunk costs. The converse is obviously not true: sunk costs by themselves will not create a barrier to entry. A casual glance around will reveal enormous amounts of irreversibly sunk costs (i.e. past investments with no resale value) with no economic relevance whatsoever.

---

[18] Certain authors identify scale economies as "innocent" or "innocuous" barriers to entry (see e.g. Salop (1979), p. 335). The distinction between "harmful" and "innocent" barriers to entry, however, is untenable. Since "barriers of entry" are constituted by single non-replicable factors of production, scale economies are the very essence of "barriers to entry".

Understanding that the notion of a barrier to entry is intrinsically linked to the existence of sunk costs, indivisibilities and scale economies, allows also taking a new look at the theory of contestable markets championed by Baumol, Panzar and Willig. The theory of contestable markets in fact implicitly makes the same argument developed on the preceding pages but states it in its reverse form: fixed costs lose their capacity to function as barriers to entry to the extent that they are neither sunk nor irreversible. While one can (and must) quibble with the notion of a fully tradable indivisibility on empirical grounds, the theory of contestable markets is fully complementary to our approach on an analytical level.[19]

Some recent examples: Ingenious constructions on fragile assumptions

It is instructive here to review three modern classics of the work on barriers to entry which have largely shaped current approaches to the issue, the contributions by Michal Spence (1977), Avinash Dixit (1980) as well as the one by Philippe Aghion and Patrick Bolton (1987). In the following we will concentrate on their treatment of the issue that interests us, the formulation of an empirically relevant and analytically meaningful characterization of barriers to entry.

One of the most influential modern contributions to the theory of entry is the article by Avinash Dixit "The Role of Investment in Entry Deterrence" (1980). The article presents itself as a formalization of the Stackelberg leader-follower model by reformulating it as a sequential choice of capacities. In addition, it situates itself in explicit contradiction to the excess capacity model developed by Michael Spence in "Entry, capacity, investment and oligopolistic pricing" (1977).

Spence in fact had postulated that firms hold excess capacity as a structural barrier to entry in order to be able to expand output at short incurring only variable cost once a concrete threat of entry has materialized. Dixit questions whether the threat of the incumbent to maintain production at the entry-deterring level is credible, i.e., whether the incumbent will maintain the unprofitable higher output level once entry has taken place. The question is whether the threat to raise output that is implicit in the barrier to entry (the excess capacity) is credible:

> "In the model used by Spence, it is simply assumed that a prospective entrant expects the established firm will respond to entry by producing an output level equal to its pre-entry capacity… It is then possible that constrained monopoly profits made by keeping capacity at the entry-deterring level… exceed what is possible with a lower capacity… When the credibility of the threat is questioned, matters can be different… (Dixit (1980), 100)."

There is indeed no guarantee in the Spence model that keeping capacity at the entry-deterring level and engaging in limit-pricing is more profitable than accommodation once entry has taken place. Hence Dixit's questioning of the credibility of such behavior seems justified.

[19] The reason why the notion of a fully tradable indivisibility is slightly extravagant is that the very notion of an indivisibility refers to the incomplete codification of its constituent parts – otherwise the different elements making up the indivisibility would be identified as separate factors of production. An indivisibility is thus always connected to a certain degree of informational complexity (see Keppler (1998)). In a market, such informational complexity, however, always translates into some degree of transaction costs. The notion of a fully tradable (or irreversible) indivisibility, which is at the heart of the theory of contestable markets, thus constitutes a conceptual horizon that is for principal reasons unattainable in practice.

It is justified indeed, when considering the formal model that Spence presents on the first pages of his article, where unspecific capacity can be expanded at an average cost equal to the rate of interest without incurring any irreversible set-up costs. In such cases, maintaining idle capacity against a mere threat of entry or maintaining production at the level of capacity once entry has taken place is indeed not credible. Remarks by Spence towards the end of the article, however, do justice to the original excess capacity story, precisely by insisting on the irreversibility of the investment:

> "The central feature of the entry deterring strategy is that the industry invests now in some form of capital, and that that investment reduces the prospects of a potential entrant and hence the probability of entry. The investment can be plant and equipment and the capital capacity. But the investment could also be in advertising and the capital customer "loyalty". The investment could also be the installations of research outlets and the capital of the distribution system (Spence (1977), 542)."

Clearly Spence is right here. To the extent that investments are irreversible and by implication firm-specific such as the building of brand loyalty or distribution channels they do constitute credible barriers to entry. To the extent that they present indivisibilities they will naturally generate increasing returns to scale and excess capacity. The problem is, of course, that it is not generally possible to study the precise measure of the power of deterrence, due to the fact that the elasticities of substitution between the goods of incumbents and entrants would need to be specified individually for each particular case.

Probably for this reason, *i.e.*, the loss of generality, Spence is loath to draw the full implications of his observations and concludes by falling behind the clarity he has just won:

> "Entry barriers are a combination of structural and technological factors on the one hand, and obstacles that are put in place by the existing industry on the other. The latter include *more or less* [our italics] irreversible investments in a variety of kinds of capital. In a homogenous product industry, a natural candidate is capacity, though that does not preclude other factors like a distributing system. In a differentiated product industry; advertising and other marketing activities that affect demand and raise the ante for the entrant also have the desired effect (Spence (1977), 542-3)."

Distribution systems and advertising activities can constitute barriers to entry to the extent that they are intrinsically linked to the incumbent firm talking about "more or less irreversible investments" blurs the issue again. Barriers to entry *only* exist to the extent that investment is irreversible and a specific distribution system foursquare implies differentiated products at the level of the end-user and monopolistic competition. Excess capacity that is not firm-specific will not, as Dixit correctly points out, act as a barrier to entry.[20] For the time being, the short

---

[20] Spence goes on to draw the implications of irreversibility and to establish its like to product differentiation, albeit without committing himself fully to this vision as being the only tenable interpretation of barriers to entry:

> The irreversibility of the investment is important for two reasons. One is that it is a way for the existing industry to commit itself in advance, a way to issue a credible threat. Secondly, there is no need suboptimally to set a relatively flexible instrument like the price, since that can be adjusted within the time horizon required for entry to take place. Moreover, with differentiated products, it is not clear that prices could deter entry even if they were inflexible. Certainly as compared with the homogenous case, the power of price to entry is attenuated (Spence (1977), 544).

The last two remarks make reference to the undetermined nature of the combined demand of two differentiated goods whose elasticities of substitution are not defined.

story in the existing literature of barriers to entry as excess capacity ended here, all the while being the only consistent story as we argue on these pages.

While Dixit's questioning of the credibility of threat generated by unspecified excess capacity in Spence's specific model is justified, his own article does not offer a satisfying alternative. In Dixit (1980), the two competitors produce undifferentiated goods essentially under constant returns to scale. While fixed set-up costs are mentioned, they play no economic role in the sense that they have no influence on either output, cost or prices. Output equals capacity, which is independent of the mentioned fixed set-up costs. Consequently, in his rendering of the Dixit-model in *The Theory of Industrial Organization* (Tirole (1989), 314-317) does away with the fixed costs, setting total investment equal to capacity and both equal to output.[21]

The limitations of the Dixit model to provide a convincing representation of barriers to entry are even more obvious than in the case of Spence. In the absence of product differentiation and third-party entry, tacit or explicit collusion would clearly be a superior strategy to the first-mover choice of a profit-maximizing output given the competitor's reaction curve. The real issue, however, is the absence of any deterrent for third-party entry. While Dixit specifies initially the existence of fixed costs, these fixed costs are devoid of any indivisibilities. They are fully malleable and adjustable and are thus unable to act as barriers to entry. In other words they are not truly sunk. Additional competitors should thus be able to enter the industry, which would result in the competitive outcome with prices equal to marginal and average costs.[22]

A similar argument applies in only slightly different form to one of the most celebrated and influential articles of the past twenty years on this issue, "Contracts as Barriers to Entry" by Philippe Aghion and Patrick Bolton. In their case, the intrinsic, non-replicable factor comes in the form of a long-term contract that cannot be resold on the market. The elegance of their model has lent great credibility to their key result that long-term contracts constitute a barrier to entry with important repercussions for anti-trust policy.

However, in the model of Aghion and Bolton long-term contracts function as barriers to more efficient entrants only because of the uncertainty about an entrants' production function.

---

21 Tirole actually calls his model the Stackelberg-Spence-Dixit model. To the extent, however, that there is no excess capacity in his model, it owns little to the former two and boils down to a simplified version of the Dixit-model and, in fact, arrives at identical results. Both Dixit and Tirole also make explicit references to the Stackelberg model (Stackelberg (1934). This is fallacious on two accounts. First, Stackelberg specified, contrary to both Dixit and Tirole, that both his competitors produced under increasing returns to scale with firm-specific fixed costs. Given that he continues to work with undifferentiated goods, however, Stackelberg is not exempt from the criticism that joint profit maximization would be a superior strategy under the point of view of static profit maximization. Second, the operative barrier to entry in the original Stackelberg model, however, is a competitor's risk appetite to engage in what is known as "war of attrition", i.e., both competitors acting as "leaders" putting out their "independent" output. Stackelberg's problem was that both competitors were likely to engage in the sub-optimal strategy of a Darwinian struggle for domination and consequently advocated forced cartelization as a means to reduce social welfare costs.

22 A case where entry is impossible but production still takes place under constant returns to scale is best exemplified in the canonical example by Cournot (1838), where the two competitors each own a natural spring producing mineral water (a non replicable asset) at zero cost. New entry is impossible even though production takes place at average cost.

Behind the treacherous generality of the term "contract" hide highly specific, idiosyncratic contractual agreements between two individual companies that depends on informational asymmetries and, presumably, high transaction costs to overcome them.[23]

Many long-term contracts, especially in the energy sector, where "Contracts as Barriers to Entry" was very influential, instead concern the delivery of standardized, highly tradable commodities. It is not so much the existence of long-term contracts *per se* that constitutes a barrier to entry but the fact that the information contained in them is private. If contractors were forced to publish the terms of their contracts and if, in addition, contracts displayed a certain degree of standardization in order to reduce transaction costs, such contracts would loose their power to act as deterrents since more efficient suppliers can always offer more attractive terms than the contractors can generate on their own. In the absence of transaction costs, this would enhance private as well as social benefits. [24]

By "externalizing" the idiosyncratic factor and the concomitant commitment value in form of an inalterable intra-firm relationship, Aghion and Bolton attempt to save the theory of "barriers to entry" once more for a theory of undifferentiated goods. Strictly speaking, the informational asymmetries and the transaction cost to overcome them imply that we are again in a world of differentiated goods and monopolistic competition.

The issue is of great real world importance. Anti-trust regulators, most notably in Europe, concerned with maximizing social welfare focus currently exclusively on long-term contracts with a view to suppressing them wherever politically feasible with scant consideration for their widely acknowledged benefits. A more enlightened approach would concentrate on the opacity and privacy of the information contained in such contracts with a view to reducing asymmetrical asymmetries and transaction costs which are the true, in fact the only, barriers to entry in the case of long-term contracts for the supply of homogenous commodities.

In the end, the literature of the past 30 years has not been able to solve the underlying contradiction of the research on barriers to entry: it is impossible to deliver a coherent theory of the issue that works with undifferentiated works. A barrier to entry is operative only to the extent that is an inalienable, non tradable and non replicable element of the production function that differentiates the output of the incumbent from that of the entrant. In this case, however, the observation by Aghion and Bolton applies that:

> "Entry may be a very complicated event to describe, when a firm can enter with a non-homogenous good… the event of entry may be difficult to observe, let alone to verify (Aghion and Bolton (1987), p. 393)."

Barriers to entry will remain "barriers to understanding" as long as researchers persist as viewing them as generalized phenomena independent of the decisions, technological processes and products of the firm producing them.

---

[23] In the world of Aghion and Bolton's article the long-term contract between a distributor and a monopoly supplier freezes out moderately more efficient suppliers, by offering a very low price which only highly efficient entrants can accept, thus partially capturing the rent resulting from the entrant's cost advantage. The two contractors know only a probability distribution of the entrant's production costs and the entrant does not know the form of the contract. If there was common knowledge of both, the two incumbents could either appropriate the entire surplus from entrants' lower costs or moderately more efficient suppliers could by out the contract and thus enter the market.

[24] This confirms, of course, once more Coase's insight that "…with zero transaction costs, the value of production would be maximized (Coase (1988), 158)."

5. Conclusions

This article aims at clarifying the concept of a "barrier to entry". It argues that past research on the issue – despite its many interesting insights – has been marred by a fundamental ambiguity. This ambiguity is due to the attempt to analyze barriers of entry in the context of competition between identical firms in industries with perfectly homogenous goods. We show that the notion of a barrier to entry has economic pertinence only if a barrier to entry is interpreted as a "unique, indivisible factor of production for which no market exists" which is possessed by a specific firm and which provides a competitive advantage over actual or potential rivals either by being able to produce at lower costs or by being able to offer more attractive products. In short, the concept of a barrier to entry only makes sense in the context of a framework of differentiated goods and monopolistic competition.

Scale economies and sunk cost, which are both frequently – and rightly – linked to barriers to entry, in fact, confirm this point of view. Early contributions by Joe Bain skirted the issue by approaching the phenomenon from an *ex post* point of view, in the sense that barriers to entries exist where profit margins are higher than average. George Stigler's contribution essentially foreshadows our conclusions but falls short of drawing the full implications of its insights.

Major papers on the issue by Michael Spence, Avinash Dixit or Philippe Aghion and Patrick Bolton have greatly increased our understanding of the mechanics of excess capacity, strategic quantity games or contracts. When faced with a choice between analytical coherence and generality, all of them, however, come down on the side of generality. In none of the three modern papers discussed, the concept of barrier to entry developed stands up to closer scrutiny. Due to the tradable nature of the barrier, i.e. the fact that it is not intrinsically linked to the production process and the product, it could always be bought out by a more efficient entrant and thus fails to operate as a barrier.

These conceptual considerations link to a point of great importance for public policy making. Anti-trust regulators today see "barriers of entry" (the realized competitive advantage of an incumbent) entirely as a negative phenomenon. Following the lead of the theorists, they are steeped in notions of competitive markets and fail to see that "barriers to entry" can very well be welfare-enhancing improvements of the production process. The question is how can static losses in social welfare due to a "barrier of entry" (the possession of a unique factor of production) be balanced with the social gains resulting from the search for and development of such barriers to entry.

This does not imply a naïve hands-off approach or the assumption that markets automatically reach the welfare-optimizing constellation. It does however imply a focus not on the removal of barriers to entry but a focus on their commoditisation. The issue of long-term contracts is a good example. Transparent, standardized and tradable long-term contracts do not pose barriers to entry. Opaque, complex and illiquid long-term contracts do.

Economists refusing to abandon the identical firms, homogenous goods paradigm will ultimately need to admit that there is no such thing as a barrier to entry. More than 80 years

ago, Piero Sraffa had by and large come to the same conclusion when analyzing the compatibility of competition and increasing returns to scale:

> "From this point of view, which constitutes only a first approximation to reality, we must thus admit that commodities are generally produced under constant returns (Sraffa (1925, 1975) 49)."

For those economists who continue to believe that "barriers to entry" exist and that they are a relevant economic phenomenon remains only the patient and modest study of the specifics of particular and transient market structures defined by monopolistic competition in a dynalic perspective. In addition, they will bring to the study an open mind about the welfare impacts of said barriers and more creative approach to their regulation.

REFERENCES

Aghion, Philippe and Bolton, Patrick. "Contracts as Barriers to Entry." *American Economic Review*, 1987, 77 (3), pp. 388-401.

Bain, Joe. *Barriers to new competition: Their character and consequences in manufacturing industries*. Cambridge, MA: Harvard University Press, 1956.

Carlton, Denis W. "Why Barriers to Entry are Barriers to Understanding." *American Economic Review*, May 2004, 94 (2), pp. 466–470.

Coase, Ronald H. *The firm, the market and the law*. Chicago, IL: University of Chicago Press, 1988.

Demsetz, Harold. "Barriers to Entry." *American Economic Review*, March 1982, 72 (1), pp. 47–57.

Dixit, Avinash. "The Role of Investment in Entry-Deterrence." *Economic Journal*, 1980, 90, pp. 95-106.

Keppler, Jan H. "La quête d'une théorie générale de la concurrence avec pertinence empirique." *Revue économique,* March 2004, 55(3), pp. 557-567.

Keppler, Jan H. "Externalities, Fixed Costs and Information." *Kyklos*, October 1998, 52 (4), pp. 547-563.

Keppler, Jan H. "The Genesis of Positive Economics and the Rejection of Monopolistic Competition Theory: A Methodological Debate", *Cambridge Journal of Economics,* May 1998, 22(3) pp. 261-276.

McAfee, Preston, Mialon, Hugo and Williams, Michael. "What Is a Barrier to Entry?" *American Economic Review*, May 2004, 94 (2), pp. 461–465.

Milgrom, Paul and Roberts, John. "Limit Pricing and Entry under Incomplete Information." *Econometrica*, 1982, 50(2), pp. 443-460.

Posner, Richard. *Antitrust law: An economic perspective*. Chicago, IL: University of Chicago Press, 2001.

Salop, Steven C. "Strategic Entry Deterrence." *American Economic Review*, 1979, 69(2), pp. 335-338.

Samuelson, Paul and Nordhaus, William. *Economics*, 12th edition. New York, NY: McGraw-Hill, 1985.

Schmalensee, Richard. "Sunk Costs and Antitrust Barriers to Entry." *American Economic Review*, May 2004, 94 (2), pp. 471–475.

Spence, Michael. "Entry, Capacity, Investment and Oligopolistic Pricing." *The Bell Journal of Economics*, 1977, 8 (2), pp. 534-44.

Sraffa, Piero. “Sulle relazioni fra costo e quantità prodotta.” *Annali di Economia*, 1925, 11, pp. 277-288. French translation in Piero Sraffa, *Ecrits d’économie politique*, 1-49. Paris: Economica.

Stigler, George. “Barriers to Entry, Economies of Scale, and Firm Size.” In *The organization of industry*. Homewood, IL: Irwin, 1968, pp. 67–70.

Tirole, Jean. *The theory of industrial organization*. Cambridge, MA: MIT Press, 1989.