

NBER WORKING PAPER SERIES

WHY BARRIERS TO ENTRY
ARE BARRIERS TO UNDERSTANDING

Dennis W. Carlton

Working Paper 10577
<http://www.nber.org/papers/w10577>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2004

The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

©2004 by Dennis W. Carlton. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Why Barriers to Entry are Barriers to Understanding

Dennis W. Carlton

NBER Working Paper No. 10577

June 2004

JEL No. L0, L1, L4

ABSTRACT

This paper discusses and criticizes the usual definitions of barriers to entry. The failure of the concept of barrier to entry to incorporate a time dimension means that it is a concept that is in need of additional embellishment in order to be useful in a practical problem or for antitrust or regulatory proceedings.

Dennis W. Carlton

Graduate School of Business

The University of Chicago

1101 East 58th Street

Chicago, IL 60637

and NBER

dennis.carlton@gsb.uchicago.edu

Why Barriers to Entry are Barriers to Understanding

By Dennis W. Carlton*

The concept of barriers to entry has been a barrier to economists' understanding of industrial structure and has misled courts and regulatory agencies repeatedly as they attempt to use the concept in antitrust cases or regulatory proceedings. There are two primary reasons why it has proved so confusing a concept. First, the theoretical underpinnings for the concept arise from the structure-conduct-performance literature which itself has been shown to suffer severe theoretical problems. Second, a large part of the confusion has arisen because authors are often unclear about the precise consequences of a "barrier to entry". For example, is price too high, profit too high, entry too slow, or social welfare too low, or all of these? If the point of defining barriers to entry is to identify some (exogenous) conditions that imply social harm, one should not define "barriers" as conditions that cause social harm, unless one can identify the conditions ex ante before solving for the equilibrium. Otherwise such a definition serves little purpose.

The work of George Stigler (1968) and Harold Demsetz (1968) clarified much of the confusion surrounding the meaning of entry barriers. However, that work, like most work in industrial organization, ignored dynamics and, in particular, adjustment costs. There is often a confusion of adjustment costs with a barrier to entry, even when competition prevails. More generally, industrial organization economists have tended to ignore adjustment costs and therefore think in terms of only short run and long run, useful pedagogical tools but ones that are often inadequate to address practical antitrust and regulatory problems.

Although there is undoubtedly disagreement among economists as to what constitutes a barrier to entry, that disagreement does not always lead to great misunderstanding among economists as to the predictions of industry outcomes. By that I mean that economists are much better at figuring out the market equilibrium than agreeing on and applying definitions. But definitions and their application can have enormous effects if they are used, as here, as a tool of analysis in antitrust cases and in regulatory proceedings. Then, disagreement over whether a barrier exists or not can matter in terms of the outcome of an antitrust trial or regulatory proceeding, even if there may not be disagreement among economists about a market's equilibrium.

Section I reviews the basic theoretical premise underlying Bain's work on barriers to entry and shows using recent work why that theory is not right. Section II explains the clarifying contributions of Stigler and Demsetz.

Section III explains why dynamics are a much more interesting phenomenon than entry barriers from both a theoretical and practical viewpoint. It discusses the consequence of failing to consider dynamics and the confusion between adjustment costs and barriers to entry. Finally, Section IV explains how courts and regulatory agencies should treat entry conditions in their analysis of antitrust cases and regulatory proceedings.

I. Bain's Barriers to Entry

Joe Bain (1956) deserves credit for trying to find presumably exogenous factors of industry structure that influence how competition occurs and that prevent price from reaching the competitive level. Bain's investigations led him to identify several factors as barriers to entry such as scale economies, large capital requirements, product differentiation, and cost advantage. These barriers protect a firm from entry and thereby enable it to enjoy above normal rates of return.

The problem with Bain's analysis of entry barriers is not his definition – namely entry conditions allowing for an elevated long run price – but his failure to articulate a consistent theory whereby the factors he identifies as entry barriers such as scale economies, large capital requirements and product differentiation lead to such an elevated price. Bain's analysis is based on a view of the world in which “barriers” determine the number of firms which, in turn, determines the competitiveness of the industry, and thereby determines each firm's rate of return. This structure-conduct-performance view of the world is, alas, too simple. The number of firms is typically determined by a decision to enter based on profitability. But profitability for any given number of firms is determined not just by exogenous factors such as costs but also by price which will be determined by the “vigor of competition” or, in game theory terms, by the competitive game being played. There is simply no reason to assume that this game is the same across different industries. As John Sutton's path breaking work (1991, 1998) convincingly demonstrates, the implication is that industries where competition is very “vigorous” will be more highly concentrated than those where competition is not as vigorous.¹ High concentration, far from being an indicator of a lack of competition, can indicate precisely the reverse!

Sutton's point can be easily seen by considering the following. Imagine an industry that in one country, Country A, is described by Cournot competition, but in another country, Country B, is cartelized with free entry into the cartel. As long as there is a fixed cost to entry, the price in Country B will exceed that in Country A, yet concentration in Country A will exceed that in Country B.

Unless one would include the "vigor of competition" in the list of factors defining entry barriers, it is just misleading to treat the number of firms as determined by "entry barriers" and it seems an odd use of language to term "vigor of competition" as an entry barrier. Indeed, the example illustrates the difficulty with treating the number of firms as determined by entry barriers alone.

I have made the point that the number of firms is determined by more than just the factors Bain claimed. But one can go further, as Sutton has, and claim that several of the factors that Bain treats as determinants of the industry equilibrium are themselves not exogenous and can be altered by investment. So, for example, firms can compete against each other by investing in the development of new products, in the promotion of the product, or in the reduction of costs. Indeed, competition along non-price dimensions can explain why as industries grow, concentration does not necessarily change but instead product quality (or advertising) increases, or costs fall. The significance of this point cannot be overstated. Models that focus on only price competition may fail miserably to correctly predict industry concentration and consumer welfare when there are other product dimensions along which competition occurs. This is likely to be particularly true in industries requiring investment and creation of new products. Indeed, I think that it is no coincidence that many of the most controversial antitrust and regulatory cases have arisen in high technology industries (e.g., computers, telecommunications) where competition in R&D and new products is paramount.

So, although we owe a debt to Bain for identifying what are interesting facts about an industry, these facts are not necessarily exogenous and do not alone determine the number of firms, which in turn does not alone determine price. Bain's quest to identify "barriers" lacks theoretical rigor.

II. Stigler and Demsetz to the Rescue

The confusion surrounding “barriers to entry” often results because the precise consequence of having an entry barrier is unclear. If there are such “barriers”, are rates of return too high? Is the existence of such barriers socially undesirable and therefore is it proper to use antitrust laws (or legislation) to attack the problem? Although Bain was trying primarily to answer the first question, many have focused on the second. I will soon explain why the two questions are very different and why the first question is not the right one for antitrust or regulation to focus on. I will also explain why the concept of entry barrier may not be particularly appropriate for answering the second question.

Stigler (1968) clarified matters by defining an entry barrier as a cost advantage that an incumbent firm enjoys compared to entrants.² With such an advantage, the incumbent firm can permanently elevate its price above its costs and thereby earn a supracompetitive return.³ This means that if the incumbent firm had spent one million dollars to establish its reputation (and had to continue to spend to maintain it), then as long as a new entrant could do the same, there is no reason to expect the incumbent firm to earn supracompetitive returns. Stigler pays no attention to dynamics or sunk costs in his discussion and focuses implicitly only on the long run steady state.

Demsetz’s classic article (1968) further clarified matters by considering a model in which it is efficient to have only one firm, an extreme example of scale economies. As long as an entrant and the incumbent are on equal footing to bid for customers, there is no reason to expect the winner to earn supracompetitive returns. Demsetz’s analysis, like Stigler’s, pinpoints symmetry as the key to answering the question: what determines whether a firm can earn excess returns? Demsetz’s work provided a foundation for contestability (William Baumol et al. (1982)) in which costless entry and exit leaves all firms in a symmetric position.

Although Stigler’s definition of “barrier” as a differential cost is concise and unambiguous, it does raise the question of why call it a “barrier”. Why not call it “differential cost advantage”? This may seem overly pedantic, but introduction of unnatural use of language can lead to confusion. Consider for example an industry where the government restricts the number of firms to 100. It issues 100 licenses to operate that are then sold in an open

market. We know the entry restriction is likely to be inefficient, but as long as all firm have access to the (artificially) scarce license at the market clearing price, there is no entry barrier according to Stigler's definition. All firms earn a normal rate of return. Yet there is a restriction to entry. It seems to mangle the English language to refuse to call this entry restriction a "barrier to entry". Using language in an unnatural way invites confusion in antitrust and regulatory proceedings.

III. Dynamics Should be the Focus of Attention but Barriers to Entry Ignore Them

The usual discussions of barriers to entry typically focus on the long run and ignore adjustment costs. In the short run, the concept of an entry barrier is not meaningful (since entry is by assumption not possible). But why is the long run of interest? Only because economists often slip into ignoring dynamics and go back to our simple models of short and long run. But as a practical matter, the long run may be of no interest whatsoever. It may take so long to get there that the persistence of supracompetitive profits until then turns out to be the fact of practical importance, not that these excess profits are eliminated in some far off future year.

Suppose we introduce the notion of adjustment costs. Those adjustment costs together with industry characteristics (including the competitive game) will influence the speed with which equilibrium adjusts over time. It is not typically a helpful thought experiment for public policy makers to ask what would occur if adjustment costs were zero. That is a bit like asking: if wages were zero, what would the new equilibrium be? Because there are adjustment costs, and because they are not a market imperfection (anymore than a wage is a market imperfection), one is likely to obtain misleading insights into policy by ignoring adjustment costs. I see no reason to call adjustment costs an entry barrier. Trying to use "barriers to entry" to cover both timing in reaching a new equilibrium and long run excess return is confusing.

Once dynamics enter the picture, we know that there can be all sorts of strategic behavior that advantage one firm over another. The source of any successful strategic behavior must ultimately be traceable to an asymmetry among firms. What game theory (and the contestability literature) makes clear is that dynamics allow credible commitments to be made when there are sunk costs. An asymmetry can arise because one firm is in a

market before another firm and can therefore act to make binding commitments before others. So, for example, building a plant with a large capacity in advance of others may be a way to make a credible commitment to produce large outputs, and this investment may advantage the firm making the investment.

By focusing on dynamics, we can now ask different, and I think more interesting, questions than those suggested by prior thinking dominated by concepts of the short and long run. We can ask not only whether price will eventually equal the competitive level, but we can also ask how long it will take before price reaches the competitive level. In response to say a merger that winds up raising price by 10%, how much of that price increase will be eroded by entry in two years or five years? If an incumbent firm uses exclusive 5-year contracts with its distributors and 20% of them expire each year, how long will it take for price to adjust in response to a surge in demand, compared to the case of 3-year contracts? How will uncertainty affect the option value of entering? I contend that these are much more interesting questions to answer and are of more practical import than the one implicitly posed by Bain or Stigler in defining entry barriers (i.e., what conditions allow one firm to earn long run excess returns).

IV. The Use of Entry Barriers in Antitrust and Regulatory Proceedings

Entry barriers are frequently an issue in antitrust cases and regulatory proceedings. Aside from the imprecision in its meaning, a problem with using the concept is that entry barriers are concerned with the long run, yet the long run may not be relevant for antitrust or regulatory proceedings. What matters for antitrust and regulation is not what might happen in some year far off in the future but what will actually happen now and in the near future. Rather than focusing on whether an “entry barrier” exists according to some definition, analysts should explain how the industry will behave over the next several years. That will force them to pay attention to dynamics and adjustment costs, the importance of which are recognized by some (e.g., Richard Posner (2001) and the other papers in this session).

The Horizontal Merger Guidelines of the Department of Justice and Federal Trade Commission do a good job of explaining that entry matters in merger analysis only when it is timely (e.g., within two years) and of

sufficient magnitude to keep price from rising above current levels. One may quibble about the “timely” definition but the point is clear. What should matter to policy makers is how fast entry erodes any price increase caused by a merger and not whether it eventually does so. In litigated cases (e.g., Section 2 cases), emphasis should be placed on how long entry will take before it erodes any temporary market power created by some attacked practice and whether the attacked practice creates efficiency benefits. The “rule of reason” is the correct approach here in which the costs and benefits are compared, but it often seems that possible efficiencies do not always receive full consideration. There seems to be a negative connotation to the word “entry barrier” and no recognition that without some “entry barrier” there may be no incentive to create new products or services. Indeed, as Demsetz (1982) has observed, property rights could be defined as the ultimate barrier to entry.

In some regulatory proceedings (e.g., telecommunications and railroads), there has been a tendency to rely on contestability theory as a guide to setting price. Contestability theory is often described as a theory in which there are no barriers to entry or exit, so that instantaneous entry or exit is possible. I have already explained why it is a mistake to confuse barriers to entry with factors affecting the timing of entry. They are two distinct concepts. But contestability theory, as commonly implemented, ignores adjustment costs.⁴ The long run equilibrium, in the absence of adjustment costs, is not usually the same as the long run equilibrium with adjustment costs for growing industries. Where the two equilibria differ, one obtains misleading policy advice by ignoring adjustment costs. In other words, it is incorrect to base policy on models with no adjustment costs on the grounds that those are the models most likely to reveal the efficient pricing since in those models there are no barriers to entry.

V. Conclusions

The words that one uses often can have unintended consequences when their meaning is unclear or even when their meaning is clear to the speaker but not to the listener. Barriers to entry, as identified by Bain, is a confusing concept. Barriers to entry as defined by Stigler is clear, but perhaps strange because the words mean something other than what would naturally come to mind. In any case, the failure of the concept of barriers to entry

to incorporate a time dimension means that it is a concept in need of additional embellishment in order to be useful in a practical problem or for antitrust or regulatory proceedings.

REFERENCES

- Bain, Joe.** *Barriers to New Competition: Their Character and Consequences in Manufacturing Industries.*
Cambridge, MA: Harvard University Press, 1956.
- Baumol, William; Panzar, John and Willig, Robert.** *Contestable Markets and the Theory of Industrial Structure.*
New York, NY: Harcourt, Brace, Jovanovich, 1982.
- Carlton, Dennis and Perloff, Jeffrey.** *Modern Industrial Organization.* Reading, MA: Addison Wesley, 2000.
- Demsetz, Harold.** "Why Regulate Utilities." *Journal of Law and Economics*, April 1968, 11, pp. 55-65.
- Demsetz, Harold.** "Barriers to Entry." *American Economic Review*, March 1982, 72(1), pp. 47-57.
- Posner, Richard.** *Antitrust Law: An Economic Perspective.* Chicago, IL: University of Chicago Press, 2001.
- Stigler, George.** *The Organization of Industry.* Homewood, IL: Richard D. Irwin, 1968.
- Sutton, John.** *Sunk Costs and Market Structure.* Cambridge, MA: MIT Press, 1991.
- Sutton, John.** *Technology and Market Structure.* Cambridge, MA: MIT Press, 1998.
- Weitzman, Martin.** "Contestable Markets: An Uprising in the Theory of Industrial Structure: Comment."
American Economic Review, June 1983, 73(3), pp. 486-7.

* Graduate School of Business, University of Chicago, Chicago, IL 60637 (dennis.carlton@gsb.uchicago.edu). I thank Robert Gertner, Gregory Pelnar, Sam Peltzman, Richard Posner, Hal Sider and the other session participants for helpful comments.

¹ The “vigor of competition” can be precisely defined. Industry A is more vigorously competitive than Industry B if, all else equal, price is lower for Industry A for any given number of firms. See Sutton (1991).

² The definition of a long run barrier to entry in Carlton and Perloff (2000) is a slight variant of Stigler’s.

³ This advantage, if it were say managerial skill, could be described as a rent, but in order to determine (static) efficiency for the products being produced, one must be careful to figure out whether price equals marginal cost for the advantaged firm.

⁴ Martin Weitzman (1983) proved that in continuous time contestability theory simplifies to constant returns to scale -- that is, a model with no adjustment costs.