X-Inefficiency Xists: Reply to an Xorcist

Harvey Leibenstein


Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28197803%2968%3A1%3C203%3AXXRTAX%3E2.0.CO%3B2-Y


Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/aea.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
X-Inefficiency Xists—Reply to an Xorcist

By Harvey Leibenstein*

Under the title "The Xistence of X-Efficiency," George J. Stigler (1976) wrote a critique of X-efficiency theory, indicated his distaste for the concept, and urged economists to abandon the idea. I will argue that this exercise in exorcism is just that. It achieves some of its results by unusual redefinitions of ordinary concepts. In the end it makes nonscientific appeals. Hence, this plea for exorcism should be ignored. At the same time, I am grateful to Stigler. As a byproduct of his attack he has raised some points that others have raised orally. This provides an opportunity to clarify some issues.

Stigler makes two essential points: 1) X-inefficiency is an illusion, it is really a matter of ignorance and mistakes; and 2) firms minimize costs, despite the fact that they choose different production techniques under similar circumstances. My contention is that X-inefficiency exists. It results from incomplete contracts, effort discretion, and nonmaximizing behavior, rather than lack of information or errors. These basic aspects are developed in Section I below, while Section II considers the problem of choice of technique. Stigler's paper also contains other points with which I disagree. Some of these are handled in Section III.

I. The General Framework for X-Efficiency Theory

It is to be noted that under Stigler's approach, as well as that of traditional theory, X-inefficiency as a phenomenon is simply assumed away. Since neoclassical theory assumes that costs are minimized, X-inefficiency cannot exist. But of course this does not imply that X-inefficiency does not exist in the real world. Hence, to capture the real world phenomenon we must consider a larger theoretical framework than the traditional one. In several papers (1969, 1973, 1975) probably available to Stigler, and in a recent book (1976), I developed such a framework. I shall refer to this larger framework as general X-efficiency theory. Table I contrasts the neoclassical model and general X-efficiency theory. Under the latter the neoclassical model can be included as a special case.

With the aid of the concepts developed below I shall try to show that there is nothing in the operation of an economy that is inconsistent with the existence of X-inefficiency, nor does competition necessarily lead to its elimination. Space constraints permit me to do little more than suggest the nature of the basic arguments.

Rational behavior involves complete concern for both the constraints and the opportunities within an economic context. I refer to both of these concerns, that is, constraints and opportunities, as constraint concern. Complete constraint concern is the same as maximization. Selective rationality usually involves less than complete constraint concern. Also, there is a tradeoff between less constraint concern and more internalized pressure that an individual feels as a consequence of less concern. Thus an individual's personality will determine the combination of degree of constraint concern and pressure he or she would like to choose—one that he feels most comfortable with. In general the individual strikes a compromise between the way he would like to behave (very low constraint concern) and the way he feels he ought to behave, which depends on internalized standards for performance and external pressures. This implies that individuals do not necessarily or usually pursue gains to be obtained from an opportunity to a maximum degree or marshal information to an optimal degree; also, maximizing behavior is a special case in this system. The specific compromise that an individual makes between the competing demands of his id (i.e., unconstrained desires) and his superego (i.e., standard of perfor-

*Harvard University.
TABLE 1

<table>
<thead>
<tr>
<th>Postulates and Basic Variables</th>
<th>Conventional Micro Theory</th>
<th>General X-Efficiency Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Psychology</td>
<td>1. Maximization or minimization</td>
<td>1. Selective rationality</td>
</tr>
<tr>
<td>2. Firm activity contracts</td>
<td>2. Given</td>
<td>2. Incomplete</td>
</tr>
<tr>
<td>3. Units</td>
<td>3. Households and firms</td>
<td>3. Individuals</td>
</tr>
<tr>
<td>5. Interpersonal interactions</td>
<td>5. None</td>
<td>5. Some</td>
</tr>
</tbody>
</table>

Performance, on the average, may be viewed as an index of his personality. However, the pressures determined by a particular context may induce an individual to pursue gains or show greater constraint concern than he would normally find comfortable. Thus personality and context select, so to speak, the degree of rationality that will control an individual's decision-making (and performing) behavior.

Our basic decision-making units are individuals even in the case of multiperson households and firms. Further, we assume that firm members have some degree of discretion with respect to the degree of effort they put forth in their work.

Effort is a complex variable which includes such discretionary aspects as the choice of: activities (A), the pace (P) at which activities are carried out, and the quality (Q) and the time pattern (T) with which they are carried out. Within some constraints people choose APQT bundles or effort points. To interpret their jobs, individuals will usually choose a set of related effort points in order to be able to meet some differential demands on effort. Also, for each person there is an effort position that contributes to cost minimization, but this is not necessarily the one that will be selected. Since in this system the firm does not control the effort levels of the individuals, it cannot necessarily minimize costs.

Note that the deviation between the optimal levels of effort from the firm's viewpoint and the actual level that individuals are motivated to put forth determines the degree of X-inefficiency in the system.

Since motivation is extremely important in determining effort levels, we have to take into account interpersonal interactions and, especially, peer group interactions, which determine the system of approval and disapproval. In turn this influences the effort level.

A vital element in our system of analysis is the concept of inert areas—somewhat similar to inertia. Individuals are presumed to choose effort positions (a set of related effort points) to interpret their jobs. The basic idea is that once an effort position (a set of effort points or APQT bundles) exists for some time period, an individual may not shift to a new position, even though there may be a gain achieved thereby because the "inertial cost" of moving from one effort position to another is larger than the perceived gain. We must keep in mind that making (or contributing to) the price, quantity, and quality decisions in a firm also requires effort.

Some may argue that maximization is really involved, if we consider the utility of the effort position minus the inertial cost. This is an illegitimate calculation. There is a distinction between person A who maximizes (zero inertial cost) and B, who does not move to a maximum point (positive inertial cost). Clearly A behaves as a maximizer and achieves a maximum position which B does not. Inertial cost explains why B does not move to a maximum. If A achieves a maximum and B does not, both cannot be called maximizers.

1The psychological terms id and superego are used in a very general sense and are not to be identified with any psychological system of thought. Substitute words can be found.
The postulates of incomplete contracts, effort discretion, inert areas, interpersonal influences, and different principal-agent objectives, as outlined, enable us to see why monopolistic firms are likely to have higher costs than the average competitive firms. Start with a firm in isolation; i.e., a firm that is not embedded in a competitive environment. Initially we ignore those motivating elements which would arise out of a competitive environment.

The basic argument follows: Because of incomplete contracts there will exist some degree of effort discretion for firm members. Thus firm members are required to interpret their jobs, which involves the choice of an effort position. Also, different interests between principals and agents, and the existence of effort discretion, imply the possibility of the choice of nonoptimal effort positions; that is, nonoptimal from the viewpoint of cost minimization. Furthermore, the existence of inert areas is consistent with the existence of nonoptimal effort positions which persist over time. Since no one in the firm is presumed to maximize profits, no one is necessarily motivated to try to get the most output from the purchased inputs, and hence costs are not minimized. In other words, under this scheme we would expect X-inefficiency to exist.

Will some degree of competitive pressure result in the reduction or elimination of X-inefficiency? It will lead to the reduction but not necessarily to the elimination of X-inefficiency. Under competition some pressure is put on some members of the various firms whose utility is affected by differential costs and relative profits so that some tightening up (i.e., adjustments) of some effort positions results. This may occur for those who are not in inert areas. For others the pressure may alter their utility functions, which in turn will force some individuals out of their inert areas and induce them to choose more appropriate effort positions in terms of X-efficiency. Additional pressure on existing firms may result from new firms entering or from the expansion of output of existing firms. As the pressures persist, further tightening up of effort positions takes place which results in reduction of costs. But this process may stop short of the point at which X-inefficiency disappears.

In the absence of pressures to contain costs there is a general cost rising tendency. Recall that firm members choose their effort position in the light of some concern for the constraint bounds that exist at the time. However, some of these constraint bounds may be too confining for some employees, and some effort choices go beyond the constraints, which in turn results in pressure against the constraint bounds. Unless managers feel pressures which induce them to struggle against some of the effort choices, especially those that are beyond the “normal” constraint bounds, the result is likely to be rising costs. Thus we may visualize the firm as an arena in which there is a struggle between the cost rising tendencies due to loosening constraint bounds, and the cost containing activities of the management.

A competitive environment may not eliminate X-inefficiency for one of two reasons: There may be a lack of supply of the right kind of entrepreneurs. Assume that there is nothing formal that inhibits entry. Recall that the firm cannot control all costs. An element that determines whether or not there are new entrants is the cost expectations of potential entrepreneurs. Suppose that the X-inefficiency that exists is 20 percent above minimum cost. Now the supply of entrepreneurship may be such that there are no entrepreneurs who believe that they could enter and produce at a lower cost than those already in the industry.

Also, the existing firms may substitute entrepreneurial activities which reduce price competition for minimizing X-inefficiency activities. Let us refer to these as “market-sheltering” activities. They would include cartel creation, monopoly creation, price agreements, product differentiation, arrangements which inhibit entry, and so on. Even if top management were interested in maximizing profits they may nevertheless feel that it pays to substitute some market-sheltering activities for the cost containing struggle to decrease X-inefficiency. In neoclassical theory we do not consider activ-
ities which reduce competition as a substitute for cost-minimizing activities, since cost minimization is assumed to take place, but in the real world this is obviously a possibility. From a theoretical viewpoint there is no reason to exclude this type of substitution or to construct the theory in such a way that we assume the market structure is given.

Another possibility is that an equilibrium may be struck in which new firms enter, but existing firms simultaneously engage in market-sheltering activities which counterbalance the effects of new entrants so that on an overall basis some X-inefficiency persists. Of course, in this case we assume that the new entrants have no preference for cost-reducing activities over market-sheltering activities.

In the light of the previous discussion we have to reconsider what are minimum costs, and what constitutes a competitive environment. The minimum costs idea is straightforward. It is the cost level that would result if firm members attempted to interpret their jobs in such a way that they made effort choices which involved cooperation with peers, superiors, and subordinates, in such a way as to maximize their contribution to output. Obviously, given a system of effort discretion, there is no reason why this should necessarily occur.

By a competitive environment I have in mind a situation in which there are a fairly large number of firms arising as a consequence of no inhibitions to entry. Nevertheless, such an environment need not lead to a situation where there are entrepreneurs always ready and capable of entering the industry in such a way as to achieve minimum costs. This would be a special and unnatural entrepreneurial supply assumption.

Our system accepts the possibility of a state of affairs approximating the neoclassical competitive equilibrium—but only as a special case. In other words, the supply of entrepreneurs willing to operate at minimal cost and capable of doing so may be large enough to counteract the market-sheltering activities. But this is viewed only as a special case which may be approximated in some sectors of the economy. It seems almost obvious that given the elements entering into X-efficiency theory there is absolutely no compelling force for this to be the general case. If anything, actual experience and the arguments presented suggest this would not be so. Effort discretion would reduce this possibility from the viewpoint of the internal organization of the firm. Furthermore, the effort by firm executives to find shelters from the need to struggle against rising costs is an additional element consistent with the persistence of X-inefficiency.

Sometimes evolutionary arguments are put forth to suggest the likelihood that an economy would approximate a zero X-inefficiency equilibrium. The essence of such arguments is that those with minimum costs would survive in the industry, while those with above minimum costs would be forced out. At the same time new firms would enter. Among the new firms, those who would achieve minimum costs would again turn out to be the survivors. The trouble with this argument is its assumption of cost minimization and its persistence. In addition, it assumes that any firm reaching minimum cost will continue to stay there indefinitely. These are very special assumptions. Once we adopt the view that 1) there is a persistent tendency for costs to rise; 2) that in general the cost rising tendency has to be struggled against by management; and 3) that market-sheltering activities are a substitute for cost-reducing activities, the evolutionary argument no longer holds. A balance may be struck between the rising cost tendencies of firms and the struggles against rising costs so that the average costs are above the minimum. The environment is not given. It is created by the firm members who are in the industry. If they choose price-setting and product-differentiating activities which shelter firms from competition, achieving minimum cost ceases to be a survival condition.

3See Sidney Winter's reply, and the article by Edith Penrose, to arguments of this sort put forward by Armen Alchian and Milton Friedman.
II. Choice of Technique and Motivation

Stigler's view on choice of technique may be gleaned from the following quotation:

The near-universal tradition in modern economic theory is to postulate a maximum possible output from given quantities of productive inputs ... and to assert that each firm operates on this production frontier.... The merit of this conventional tradition is also its demerit: it eliminates the problem of the choice of technology.... one may lament ... the failure of Robbins and Leibenstein, and all of us in between to recognize the problem of determining which technologies will be used by each firm.... The choice is fundamentally a matter of investment in knowledge.... Leibenstein deserves credit for revising this Marshallian question, but his attention to X-inefficiency as the explanation is an act of concealment: it simply postulates the differences in technology among firms which should be explained.

[1976, pp. 214-15]

Stigler's assertion to the contrary, I believe that the ideas summarized in the previous section do in fact explain the choice of technique, and that Stigler's denial of the significance of motivation, and the postulate that every firm is on its production frontier, really conceal what actually goes on in the typical multiperson firm.

One of the main notions of X-efficiency analysis is that the technique of production cannot be determined by the choices of some major executive. When the firm hires specific capital and labor inputs it does not determine the actual technique. It is not a matter of mistakes, it is a matter of the extent to which some people can choose what other people do. We argue that effort discretion is a fact of life. Hence, the actual technique must depend on individual choice and on the motivating factors existing both within the firm and in the firm's environment. These in turn strongly influence detailed choices. Thus while the choice of technique from this viewpoint ceases to be mysterious, it is no longer consistent with the assumptions of neoclassical theory.

It is to be noted that Stigler's view of the choice of techniques differs from the standard textbook view in which competing firms have full knowledge, or equal access to knowledge, so that they minimize costs, choose the same technique, and achieve the same minimal cost level (see C.E. Ferguson). Unfortunately, the empirical evidence shows that this is not the case. Hence the theory makes an incorrect prediction. It seems to me that this is a straightforward view of the theory and its results. We should not be inclined to rationalize the deficiencies of the theory, but to face up to some of its limitations.

Stigler argues that all firms operate on the production frontier. However, different firms have different frontiers, possibly because of differences in knowledge, or entrepreneurial capacity—we do not really know. But if every firm has its own frontier, some are further out than others. It is a small step to think of the outermost frontier (or the frontier of frontiers). Hence, some firms are operating inside the outermost frontier, but this does not represent allocative inefficiency. It represents what I have called X-inefficiency. Thus, the idea of X-inefficiency may be hidden by Stigler's approach, but ultimately it is not really avoided. The alternative approach, one that seems more natural, is to say that some firms do not operate on the production frontier. As a consequence differences in costs between firms exist. In my view these cost differences are best explained by the fact that firm members make discretionary effort choices, and that these are influenced by the motivational environment within the firm and between firms.

Stigler argues that motivation is not an input, but part of the problem of choice of technology. But what is the choice of technology problem in the neoclassical sense if it is not to choose the appropriate combination of inputs? The critical question is whether motivation is or is not an input in neoclassical theory, in the real world, or in both? Now, motivation is not an explicit
input in the neoclassical model because the maximization of something can be viewed as assuming away variations in motivation. But this does not imply that motivation is not a significant variable in the real world.

Despite Stigler's sarcasm contained in his "Romans discovering America" example, there is little doubt that motivation is a significant variable in the real world. There is a large literature in industrial psychology and on business organizations which supports this view. In addition, all of us are aware of situations in which our own motivations changed for some reason, so that we were able to undertake and carry out certain tasks in a superior fashion or more vigorously than would otherwise have been the case. The empirical significance of motivation is not an issue. The real issue is whether we want to think about it in a straightforward manner for the analysis of economic problems. I would suggest that the straightforward approach is to recognize motivation as an obviously important factor in determining productivity, and that frequently we want economic models to contain motivations directly or indirectly as a variable. There may be sets of problems whose analysis is not helped by including variations in motivation. Here the neoclassical theory may be appropriate. However, there is a large set of other problems, especially those for which effort is a discretionary variable, where it makes sense to include motivation as a variable.

In multiperson firms, where effort is a discretionary variable, no one in the firm really controls all that goes on. The nature of the effort depends on the motivation that individuals bring to the firm from their background and personality. They themselves create the motivational environment and structure, since everyone contributes to the interpersonal system of approval and disapproval, which influences the reaction patterns to an individual's effort position choice.

Stigler argues that "the choice [of technique] is fundamentally a matter of investment in knowledge" (1976, p. 215). No doubt knowledge plays a role, but could it carry the burden of the entire explanation?

Do individuals and firms acquire all the knowledge it pays for them to acquire? Stigler provides no evidence that this is the case. I believe there is some evidence to the contrary. Stigler speaks of hypothetical farmers making choices in various ways. A recent paper by Kenneth Shapiro and Jurgen Müller examines the choice of technique by farmers in Kenya. It demonstrates that in reality farmers do not use as much knowledge as is available to them, not only because of differences in costs but also because they do not wish to. It is not unreasonable to presume that motivation plays a role here. Furthermore, it would be strange to argue that these real farmers are operating on their production frontier.

III. Other Points in Dispute

Early in his paper Stigler argues "Surely no person ever seeks to maximize the output of any one thing: even the single proprietor, unassisted by hired labor, does not seek to maximize the output of corn: he seeks to maximize utility, and surely other products including leisure and health, as well as corn, enter into his utility function. When more of one goal is achieved at the cost of less of another goal, the increase in output due to (say) increased effort is not an increase in 'efficiency'; it is a change in output" (p. 213, emphasis added). This argument does not really handle the issue. It avoids some basic distinctions, that between the product for sale and intrafirm outputs, and that between single person and multiperson production units.

Consider the three cases in Table 2. It is assumed that the product under consideration is the one sold on the market by a

<table>
<thead>
<tr>
<th>Case</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δ Cost</td>
<td>+20</td>
<td>+20</td>
<td>+20</td>
</tr>
<tr>
<td>Money value of compensating on-the-job utility</td>
<td>0</td>
<td>+10</td>
<td>+20</td>
</tr>
</tbody>
</table>
multiperson firm. Now suppose that some firm members choose to put forth their effort so that the cost of the product is 20 percent higher than minimum cost. In Stigler’s language, output is changed if productivity declines. However, I believe it’s more convenient to speak of the changes as an increase or decrease in on-the-job utility. In Case 1 we see that there is no compensating increase in on-the-job utility for the lower value of effort. In Case 2 we assume that there is some increase in utility but it is worth less than the decreased value of effort. In Case 3 it is assumed that the increased value of utility of the workers is exactly equal to the decreased value of the product. There is nothing in the argument put forth by Stigler to deny the possibility of Cases 1 and 2. Even if in some sense more “leisure” is produced, there is no reason for the value of leisure to be in any sense equal to or greater than the reduced value of the product. Cases 1 and 2 are clearcut cases of X-inefficiency without any counterpart increase in on-the-job utility.

It is even possible to visualize a more extreme case under which the counterpart result of producing less with given inputs can result in a lower on-the-job utility than would otherwise be the case. Suppose that two individuals who do not get along are put in leading positions in a department of a firm so that output is lower than would otherwise be the case. This leads to a lower aggregate on-the-job utility than would be the case if things were better arranged and they were in different departments. There is nothing to prevent any of these examples, or something like them, from occurring in real life. We can define things in such a way that such possibilities are assumed out of existence, but this simply uses theory to put blinders on our powers of observation. The main point is that X-inefficiency can exist without complete counterpart increases in on-the-job utility.

There is some discussion about the optimal enforcement of contracts in Stigler’s piece, but it is not clear what sort of contracts are being enforced and who does the enforcing. If the effort aspects of a contract are vague, there is not much that can be done to enforce them. What standards of performance should the firm pursue in its enforcement activities if the contract is vague? Furthermore, in a firm run by agents rather than owners, the extent of attempts at enforcement depends on the motivations of the agent-managers. Further, we are told that “...the avoidance of unpleasant tasks and the enforcement activity designed to curtail this avoidance can be carried on to the utility-maximizing degree and generate no inefficiency in producing utility” (p. 213, emphasis added). The language used suggests that the enforcement activity, etc., can be carried on to a utility-maximizing degree. Even if this could be the case, is it the relevant consideration? The question is, will it be? There is no evidence or logic presented to suggest that this will be the result. What is ignored in all this is the discretionary effort gap within which motivation will determine the actual effort level.

Stigler also argues that payments to inputs are adjusted to productivity, and hence this takes into account X-inefficiency. While this may be true in the neoclassical perfect competition model, it does not take account of deviations from perfect competition, nor does it take account of payments to firm members which help to create market sheltering conditions.

One of the empirical questions at issue is whether monopoly leads to higher costs than competition. Stigler provides no evidence to the contrary. There is some evidence on this matter. Walter Primeaux, Jr. compared the forty-nine cities with more than one electric power company with those that have single electric power companies. After taking economies of scale into account, Primeaux found that

...on the average, cost is reduced, at the mean, by 10.75% because of competition. This reflects a quantitative value of the presence of X-efficiency gained through competition; or an estimate of the loss caused by the absence of competition in a regulated environment. [p. 107]

The value of the X-efficiency hypothesis is reflected in two other studies. John Shel-
ton compared owner-operated franchise restaurants with manager-operated units. Despite the very high degree of standardization of menus, accounting systems, etc., and despite the virtual equality of sales volume, the owner-operated units averaged a profit margin of 9.5 percent but the manager-operated units averaged only 1.8 percent. T.Y. Shen studied technological diffusion as an element that influenced cost in 4,000 plants. He found that

... it is necessary to recognize the presence of a further systematic influence that also affects the change of input-output combinations of manufacturing plants over time ... We find the observed behavior pattern is better explained by the prevalence of 'X-efficiency' rather than by substitution ... This tentative finding is put to a further test by a perusal of the nature of a factor intensity change. Once more the X-efficiency hypothesis turns out to be more consistent with the data. We conclude that a technological change model based on diffusion requires the estimation and incorporation of X-efficiency. Until this step is taken, the use of the extended diffusion model for explaining growth is of dubious validity. [p. 264]

These represent some studies which show the value of X-efficiency theory over its strict neoclassical counterpart.

Since I have postulated a nonmaximizing framework in which maximization is a special case, some remarks on maximization are in order. The maximization hypothesis to be meaningful must be falsifiable. That is, if the word "maximum" means something, then the possibility of choosing a nonmaximum must also exist. Hence the tautological approach to maximization eliminates meaningfulness. In the third edition of Stigler's justly famous text (1966) he seems to argue in favor of the tautological interpretation of utility maximization. He defends it on the ground that at the very least this approach leads to accurate predictions. However, Stigler himself does not indicate what these predictions are. Consider the case of voting behavior. If people really maximized utility, no one would vote except those who felt that they had a deciding vote, and were concerned about the outcome. The evidence shows that in fact the majority of the population frequently does vote, although on the average a higher proportion will vote when the election is close or strong private interests are involved. This is just one of many examples of this kind. In other words, the maximization postulate can be false in many instances. Nevertheless, my general view is that in some cases the maximization model may be a useful simplification, but we should not attempt to use it everywhere.

In general Stigler appeals to tradition and warns against "the mighty methodological leap into the unknown that a non-maximizing theory requires" (p. 216). Both the appeal and the warning are nonscientific. After over a century of maximizing models, it is time to consider nonmaximizing approaches.

REFERENCES


