Lotta *Lemmata*: A Sour Harvest

Philip R.P. Coelho and James E. McClure

Either make the tree good, and his fruit good; or else make the tree corrupt, and his fruit corrupt: For the tree is known by its fruit. (*Matthew* 12:33*)*

I. The Uses of Economic Theory

A. The Traditions of Economics

Resources are scarce relative to human wants. The tradition of economics is to use this perspective to model specific theories whose implications are then tested with *independently* generated data. *Independently*, meaning that data were not the basis for constructing the theory, or vice versa. An example of data that is not *independent* of economic theory is in *The Historical Statistics of the United States*. The data on farm capital were constructed by the United States Department of Agriculture using depreciation models derived from economic theory. Consequently if a researcher who, unaware of its provenance, uses these data on depreciation of farm capital to assess the validity of an economic theory (that depends on depreciation data as an independent variable), then he is in extreme error; in this case the depreciation data are an artifact of economic theory, they are not independently generated.[1]

Theories that do not generate testable statements, or do not do so at a reasonable cost, have difficulty establishing themselves in economics. For example, economists generally dismiss utopian socialism that holds that its advent will usher in a humanity purged of envy, lust, covetousness, jealousy, and pride. This statement is not testable; only if "True" socialism is created will a cleansed humanity be reborn. If mankind is not cleansed, then "True" socialism has not been created. This reasoning is tautological; it cannot be challenged by evidence.

Contrast this with the assumption that firms wish to maximize profits. Because the firm does not have a physical existence, this is not an operational proposition. Nevertheless, employing this assumption, we are able to derive statements that are testable. For example, in markets where firms supply small proportions of market output, the profit-maximizing model

[1] Leamer (1978) identifies other difficulties in empirical economics when the hypotheses are altered to fit the data.

implies that firms' output increase with market prices. Data either consistent or inconsistent with the theory can be obtained at a reasonable cost, allowing implications to be compared to data obtained from the "real" (outside the boundaries of the theory) world.

The ability to assess theories with evidence separates them from tautologies and other kinds of statements. Still, assessments are judgmental, not absolutes revealed by evidence; to distinguish dogma from causal explanations we appeal to evidence beyond the confines of the internal workings of theories. All theories require some background conditions for the conduct of tests. Whether it is the Newtonian theory on the behavior of falling objects, or human behavior in the face of falling prices; "appropriate" test conditions must be established. Once these are established, we can judge whether observations are consistent or inconsistent with hypotheses. This is a relatively Popperian or "operational" approach to science.[2] Operational statements are made about how things behave in the physical world and can be assessed empirically.[3] While such statements cannot be proved "true," they must be capable of being shown as either consistent, or inconsistent with observations and/or experience. Statements that are deemed to be inconsistent with the available evidence are useless in explaining the stated conditions (i.e., they can be considered to have been operationally falsified).

Empirical tests of operational theories serve at least two useful purposes; they can: 1) cull theories whose predictions are clearly inconsistent with the relevant range of experience; and 2) delineate a set of experiences within which a theory can be applied. Theories that are not rejected as empirically false, or false over the examined ranges, are used to explain and/or predict phenomena. For example, the theory of the speed of falling objects in a vacuum is useful for bowling balls impeded by normal atmospheric frictions for relatively short distances, but is of little use for predicting the trajectory of dropped feathers in the earth's atmosphere.

---

[2]See: Popper (1934/1959).

[3]Our usage of the term operationalism is parallel to that envisioned by Donald F. Gordon, not that found in Samuelson (1965). For a critique of Samuelson's criteria for operationalism see Donald F. Gordon (1955).

The economic perspective of people pursuing their self-interest in the face of scarcity has generated a myriad of operational propositions about human choices and their consequences. For example, in economic history scholars can examine the behavior of both slaves and slave owners in the light of economic theories. Economics illuminates the slave experience even when neither slaves nor slave owners were "rational" in the sense that they were consciously maximizing some objective function. In economics scarcity and competition lead to predictions about resource allocations that approximate the results that rationality could generate.[4] These are both assumptions and testable propositions; if we assume them, then we compare net returns of investments in slave capital with the returns available in other types of investments. If we debate whether slave owners were rational, we may use the same data and compare rates of return. If slave owners were irrational, we would expect there to be no correlation between rates of return on slave capital and on other forms of capital.

B. An Alternative View of Economic Theory

Over the past five decades, economic theories appearing in the general interest professional journals have become less general and more technique focused. These trends are challenging the traditional importance of operationalism. The more prominent the journal, the more emphasis there appears to be on non-operational statements. It is increasingly common to observe these journals devoting significant attention to manipulations of sophisticated mathematical systems that have no testable implications.[5]

To quantify trends in mathematical complexity we measure the usage and frequency of the term "lemma" in publications. Lemmas (or *lemmata*) are intermediate steps in mathematical proofs. These explicitly delineated steps organize complex, and lengthy proofs into forms that

---

[4]Alchian (1951) suggests that the competition for scarce resources leads to an allocation of resources consistent with profit maximization *even if agents are acting randomly*. Along similar lines, Hirschleifer (1977) and Marshall (1964) suggest that the connection between economics and biology could represent a revolutionary shift; see Kuhn (1962) on resistance to paradigm shifts.

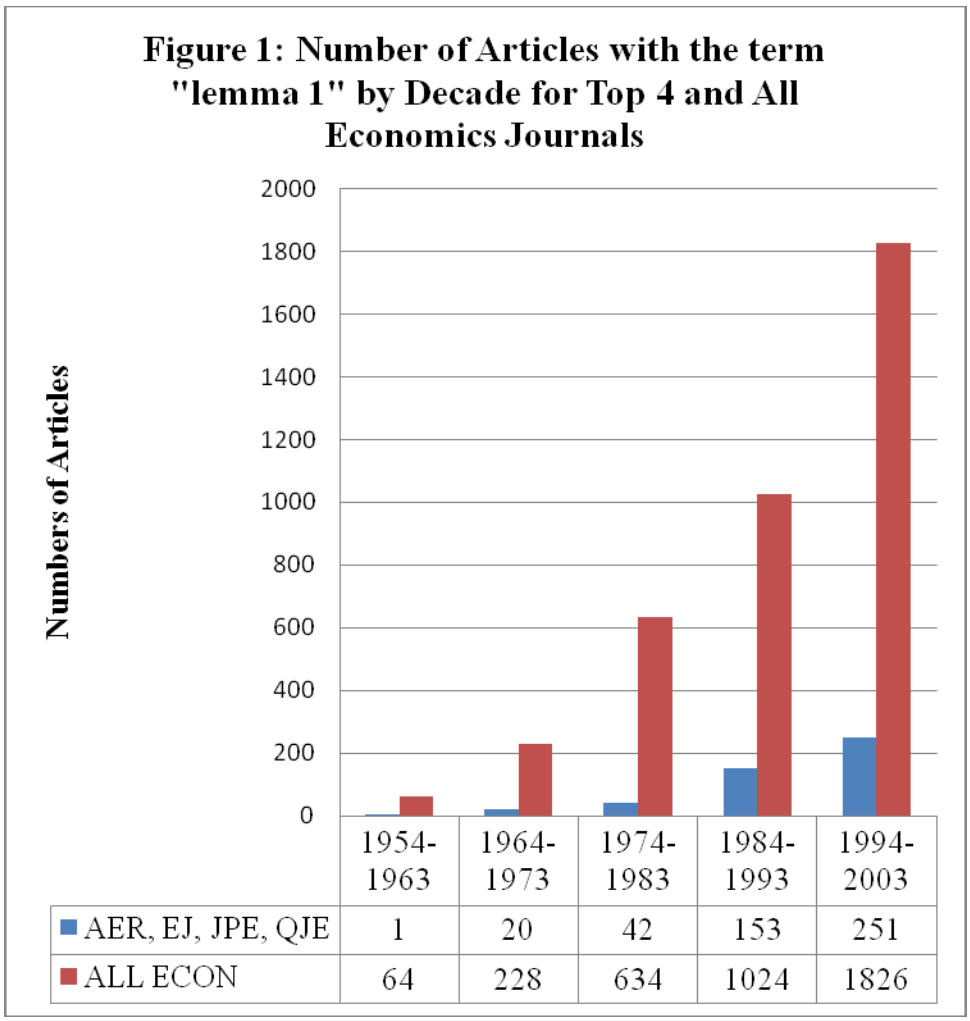
[5]The journals we are referring to are the: *American Economic Review*, *Economic Journal*, *Journal of Political Economy*, and *Quarterly Journal of Economics*.

are more easily understood. Prior to 1960, lemmas were almost never seen in general interest journals of economics.[6] Subsequently, lemma sightings have risen steeply ; Figure 1 below provides evidence on the appearance of the term "lemma 1" in four general interest journals: *American Economic Review (AER); Economic Journal (EJ); Journal of Political Economy (JPE);* and the *Quarterly Journal of Economics (QJE).*[7] The figure also provided results of the same search for all 97 economics journals in the *JSTOR* data base. The data in the histogram are in ten year increments starting from1954 through 2003:[8]

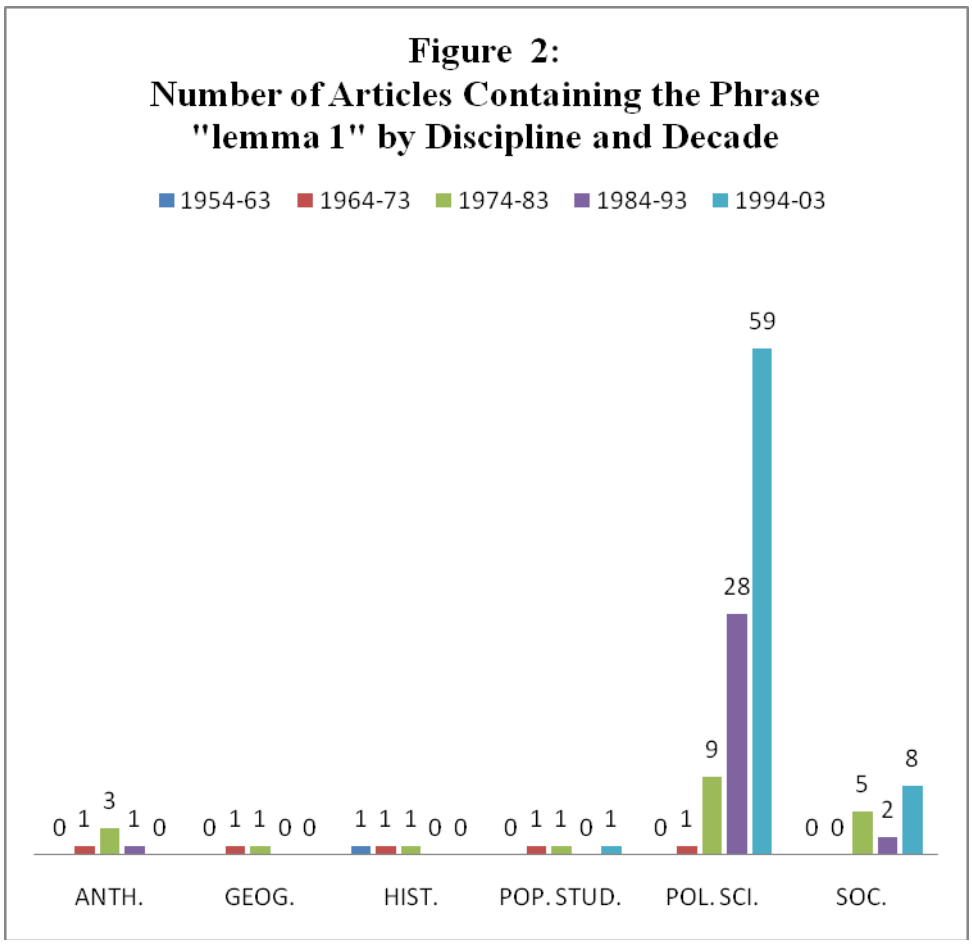
[6] Coelho and McClure (2008, footnote 7; p. 340) indicate that for the *AER*, *EJ*, *JPE* and *QJE*, the usage of lemma by decade prior to 1960 was as follows: "1900-1910 (*one article*); 1911-1920 (*two articles*); 1921-1930 (*zero*); 1931-1940 (*zero*); 1941-1950 (*two*); 1951-1960 (*one*)."

[7] The search term "lemma 1" rather than the alternatives "lemma" or "lemmas" was used because searches using the latter terms brought up too many: 1) "*di*lemma[s],"2) authors whose names had "lemma " in them, and 3) other unidentified problems. These problems were less frequent in the economics literature, but cross-disciplinary consistency demanded that our searches be identical.

[8] This extends the series shown in our 2008 article. The division of decades beginning in 1954 and ending in 2003 allowed the most current information to be included using the *JSTOR* database; in *JSTOR*, the *EJ, JPE,* and *QJE* have "firewalls" that end searches at 2003.



Beyond the discipline of economics, we conducted the same searches for ALL journals in the JSTOR data base for the following disciplines: Anthropology (47 journals), geography (18 journals), history (157 journals), population studies (23 journals), political science (71 journals), and sociology (75 journals). The data on the frequency of the usage of "lemma 1" in each of these disciplines by decade is summarized in Figure 2 below.



Comparing Figures 1 and 2, it is apparent that economics has the lion's share (almost all) of lemma usage amongst the disciplines considered.[9] The trivially small usage of complex mathematics in disciplines outside of economics does not provide sample sizes that allow cross-disciplinary analysis. In the most recent decade, compare the totals for All Economics in Figure 1, to the sum of the same decade totals for all Anthropology, all Geography, all History, all Population Studies, all Political Science and all Sociology: 1826 articles in economics versus 68 in all the other disciplines combined! The leader in lemma usage in non-economics journals is Political Science. In the most recent decade, lemma usage in all Political Science journals is 59 articles; importantly, this was a level surpassed by ALL ECONOMICS journals over half a

[9] On a normal size page, Figures 1 and 2 cannot be combined because the vertical scales are so different; trying to combine them would shrink the non-economics data into a series of almost invisible bumps.

century ago. Because of the relatively trivial usage of complex mathematics in non-economics disciplines, it is apparent that mathematical complexity is relatively unique to economics. Thus we focus this paper on the discipline of economics.

The vast majority of mathematically complex articles in economics are primarily devoted to two tasks: 1) telling a story that is consistent with some "stylized facts" within a mathematical model; and 2) deriving mathematical proofs of the internal consistency of the model. These "proofs" are distinct from evidence; they neither rely on anecdote, data, nor history. For example, a mathematical "proof" for an assertion of an upward-sloping demand curve does not mean that there is any evidence that supports the assertion.[10] The term "proof" only means there are no obvious mathematical inconsistencies.[11]

Inspecting the discussions that accompany these mathematical exercises, it appears that the criteria influential in their publications were aesthetic; the rhetoric suggests the articles are valuable because they are "elegant," "original," "imaginative," "innovative," or "suggestive." The question arises: What is the utility of elegant mathematical complexity unconstrained by ugly facts? Disciplines other than economics that rely upon empirical verification do not consider mathematical aesthetics sufficient to justify incorporation into the canon. For example, in physics explanations for the world have been repeatedly overthrown by data; from the Michelson-Morley experiments on the "ether," to quantum theory that has made the pursuit of a "grand unified theory" of physics something to be relegated to the far future, if ever. Physicists keep the quantum theory, not because it is "elegant," or "suggestive," but because it generates a series of testable propositions that have not led to its refutation, and it has a wide variety of uses

[10]W. Pesendorfer (1995) develops a theory of fashion cycles whose assumptions include: 1) the demand for design is discontinuous and sometimes upward sloping; and 2) that the costs of copying designs are a significant barrier to competing firms producing virtual copies. His complex mathematics may be internally consistent, but the real world is wildly inconsistent with these assumptions. Both editors and author appear indifferent to the wild inconsistencies with observables; see Coelho, Klein, and McClure (2004); Pesendorfer (2004); and Coelho, Klein, and McClure (2005); and Pesendorfer (2005).

[11]Gödel's Theorem on the inadequacy of mathematical systems to determine the truth or falsity of all mathematical statements is never mentioned.

in science and industry. In stark contrast are the recent trends seen in economics where the empirical content and the generation of testable hypotheses are inconsequential to the assessment of published articles.[12]

Many of the articles one currently finds in mathematical economics are *artifacts* devoid of utility in the world of commerce, prediction, or explanation. In an article (2008) we examined all the citations available at that time to the articles in the 1980 volume of the *Journal of Economic Theory* that contained 5 or more lemmas (there are 12 articles). We found that out of the 237 articles that cited the lemma-laden *JET* publications, only *two* attempted an empirical assessment, neither had hypotheses that were unambiguous.[13] The question arises: Is this specific to the *Journal of Economic Theory*? More generally: Do lemma-laden publications in top general interest journals yield a larger number of citing works containing empirical tests assessing the hypotheses in the lemma-laden source articles?[14]

Using the *JSTOR* database we searched in the *American Economic Review, Economic Journal, Journal of Political Economy*, and *Quarterly Journal of Economics* for the years 1980 through 1987 for all articles that had five or more lemmas in them. We found a total of nine articles. Table 1 summarizes our findings:

---

[12] Indifference toward observational reality and evidence has produced untoward results. Mathematical models touted by R. McAfee and J. McMillan (1996) as the "triumph" of "mathematical economics" and "game theory" were employed by the FCC in the United States in auctions of the wireless radio spectrum. The "triumph" was short-lived; auction participants gamed the system by submitting bids that exceeded expectations. When auction "winners" failed to resell the purchased spectrums, they reneged on their contracts; some avoided paying by entering bankruptcy. See A. Girard (2001) for a discussion of the outcome of these game-theory-inspired-auctions. Girard's paper (p. 90) ends by quoting from a court ruling that: ". . . the Commission violated the provision of the Bankruptcy Code that prohibits governmental entities from revoking debtors' licenses solely for failure to pay debts dischargeable in bankruptcy." If upheld in higher courts, then this is anything but a "triumph" for "mathematical economics." The failure of the "triumph" of game theoretic models has not been widely publicized in major economic journals.

[13] For more on the methodology of *JET* see: "Model building versus theorizing: The paucity of theory in the Journal of Economic Theory" by D. Klein and P. Romero (2007).

[14] We are indebted to an anonymous referee for causing us to consider this issue.

TABLE 1: All Articles with Five (or more) Lemmas in General Interest Journals, 1980-87							
	Characteristics of Citing Articles						
Source article authors (journal; year)	**Number citing articles WOS**	**Number citing articles assessed**	**Number containing data**	**Number testing a source hypothesis**	**Number accepting or rejecting a source hypothesis**	**Number with consistent empirical results**	**Number with inconsistent empirical results**
Hellwig (*AER*; 1981)	**16**	**13**	**2**	**0**	**0**	**0**	**0**
Herberg & Kemp (*AER*; 1980)	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Harris & Weiss (*JPE*; 1984)	**16**	**16**	**10**	**0**	**0**	**0**	**0**
Laffont & Tirole (*JPE*; 1986)	**298**	**291**	**52**	**4**	**0**	**3**	**1**
Malcomson (*JPE*; 1984)	**133**	**122**	**39**	**0**	**0**	**3**	**3**
Ross (*JPE*; 1987)	**30**	**29**	**5**	**0**	**0**	**0**	**0**
Azariadis (*QJE*; 1983)	**56**	**56**	**2**	**0**	**0**	**0**	**0**
Khan & Vohra (*QJE*; 1987)	**19**	**19**	**0**	**0**	**0**	**0**	**0**
Roemer (*QJE*;1986)	**54**	**51**	**0**	**0**	**0**	**0**	**0**
TOTALS	**622**	**597**	**110**	**4**	**0**	**6**	**4**

The results in Table 1 indicate that of the 597 citing articles that were reviewed, ten had tests that indirectly and directly[15] addressed the theory of the source article; consequently 1.68 percent of the citing articles had *any* hypothesis testing. The scarcity of empirical assessments of mathematically complex articles in general interest journals is similar to that found in the empirical assessment of lemma-laden articles in the *JET* source articles.

The want of empirical insights stemming from lemma-laden articles is not because authors have discouraged "future research" to provide them; quite the contrary, appeals for future research are common.[16] Two questions arise: 1) what inhibits mathematically complex economic theories from being operationalized? 2) Why have mathematically complex theories gained greater prominence in the face of their inconsequential empirical yields? With regard to the first

[15] The last two columns of Table 1 contain *all* the results that appear in the column headed "Number testing a source hypothesis." The total of ten articles in the summation rows of the last two columns includes the results of the citing articles that tested a source hypothesis.

[16] Conversely, it is not uncommon for the issue of operational testing to be ignored altogether. While this approach is less disingenuous, it contradicts scientific tradition.

question we offer two inter-related explanations: 1) in testing mathematically complex hypotheses difficulties arise because the surreal assumptions within the models make the process of looking for an appropriate test never-ending; and 2) the hypothesis of Donald Gordon (1955) that mathematical complexity inhibits operationalism. This proposition is examined more fully in Section II.

The second question is on why have mathematically complex stories have gained prominence. The answer is because journals are much less receptive of papers that present evidence (empirical, observational, or anything that depends upon real-world observations) that contravenes previously published papers. Observational reality that used to constrain complex mathematical theorizing (and the "stylized facts" it embraces) is no longer a deterrent to fantastic models because papers that comment on previously published papers are met with overt editorial hostility. The evidence for the animus against and decline in critical commentary is in Section III.

II. Mathematical Complexity versus Operationalism in Economics

We are not methodological absolutists; similar to Deirdre McCloskey (1983), we are arguing that in investigating operational issues, economists make reasonable searches for knowledge in the world, and in assessing explanations confront seemingly contradictory evidence directly. We believe that evidence that is not self-referential has a major role in the assessment of theories. This approach dates back to John Stuart Mill. Todd G. Buchholz (1990) explains:

> With seemingly Solomonic wisdom, Mill sliced out a role for each [deductive theorizing and inductive empirical analysis]. Each method could balance each other. If some economists deduced from flawed *a priori* principles, empiricists could throw observed counterexamples in their faces. (p. 97)

Evidence is taken from the real world, and to operationalize an explanation we must judge it empirically. Operationalism at work is any assessment of a statement that relies upon data or experiences that are independent of the statement and not fabricated. Lemmas, and other intermediate steps, that culminate in the mathematical proof of an economic theorem are self-referential and usually non-operational. The mathematically derived and "proven" theorems may, or may not, be operational.

Because the realms of mathematics and the world outside of it are not the same, special care has to be used in applying mathematical tools and conventions to insure operationalism. In a famous passage Alfred Marshall (1920; 1964) recognized this. While Marshall condemned long chains of mathematical reasoning in economics, he extolled the value of mathematics in training economists to grasp "mutual interactions":

> It is obvious that there is no room in economics for long trains of deductive reasoning; no economist, not even Ricardo, attempted them. . . . But a training in mathematics is helpful by giving command over a marvelously terse and exact language for expressing clearly some general relations and some short processes of economic reasoning; which can indeed be expressed in ordinary language, but not with equal sharpness of outline. And, what is of far greater importance, experience in handling physical problems by mathematical methods gives a grasp, that cannot be obtained equally well in any other way, of the mutual interaction of economic changes. (p. 644)

As noted by Samuelson (1957, p. 57), Alfred Marshall and John Stuart Mill were both given to "speaking of the dangers involved in *long* chains of logical reasoning." Discussing Marshall's view, Samuelson stated that: "Marshall treated such chains as if their truth content was subject to radioactive decay and leakage-at the end of *n* propositions only half the truth was left, at the end of a chain of *2n* propositions, only half of half the truth remained, and so forth in a geometric multiplier series converging to zero truth." Donald F. Gordon (1955, p. 58), also noted Marshall's aversion to "long chains of reasoning." The empirical orientation of Marshall was emulated by economists through the mid-twentieth century.

A. The Gordon Hypothesis

Building upon Mill and Marshall, Donald F. Gordon (1955; 1968) considered the utility of mathematics in economics. Gordon hypothesized that applying sophisticated mathematical techniques to economic phenomena was unlikely to generate empirically valid propositions. He argued that operational propositions and sophisticated mathematical modeling were in opposition because the *ceteris paribus* assumption that requires the relationships among variables to be stable is much more likely to be violated the longer the chain of relationships and variables:

> . . . the essential point is the difference between theories using a large number of functions and those using one or two, since formal and mathematical reasoning is normally required when the number of relationships simultaneously being considered becomes large. As we have seen, even though each may be quite plausible, a combination of very many [relationships] will rarely be so [stable]. Consequently, it happens that the cases in which formal and mathematical reasoning is most likely to be required are precisely the cases in which, for other reasons, the validity of any conclusions is likely to be conjectural. It is frustrating but nevertheless true that, where mathematics is most likely to be useful, the theory is least likely to be valid, while, where the theory is most likely to be true, complex deduction is generally not needed. (p. 58)

Gordon explained the operational difficulty that longer chains of mathematical functions created by using an example of a theory relating three distinct variables x, y, and z:

> Again, the relationship between x and y may be stable long enough for a shift along that function but not stable long enough for a shift along that function plus a subsequent shift along another [z]. (p. 53-4)

Problems occur whenever there is a breakdown in the *ceteris paribus* assumption. A breakdown occurs when the chain of relationships from *x* to y and then *y* to *z* unfolds in real time rather than instantaneously. Gordon reasoned that because economic phenomena are time-dependent, the longer the chain of functions that were linked together in a theory, the more likely it was that the passage of time would materially alter the specified relationships in unpredictable directions. Gordon saw the timelessness implicit in mathematical relationships as an impediment to operationalizing them; the greater the number of mathematical linkages, the greater the impediment.

Simple models are productive because they foster the development of operational propositions. But not all simplifying assumptions are productive. There are simplifications that both damage reality and inhibit the development of operational propositions. Again, Gordon hypothesized that many mathematically sophisticated economic models embodied unproductive assumptions because temporal effects were ignored, and the consequences of indirect effects occur over real time, not instantaneously.[17] In complex economic models the *ceteris paribus* assumptions implicit in the functions of a theoretical system are more likely to be false because falsity increases with the number of functions. If real-world elements of the theory materially affect either the model's causality or the magnitude of its effects, then a model's false assumptions reduce the ability to explain and/or predict. Each additional false assumption that affects the model exponentially reduces the theory's utility. When a variable is assumed to react instantaneously, but actually reacts over real time, is an example of an assumption that materially affects the model. The more time that is required to allow for the reactions and feedback effects of a specific shock to work through the system, the greater the likelihood that subsequent events will completely swamp the original shock. In a general equilibrium system, where *all* sectors and *all* economic actors are modeled, the probability that other events will occur in real time before a specified shock has even partially worked its way through the model approaches unity.[18]

Alternatively we can analogize Gordon's proposition with the theory explaining the acceleration of falling objects. For most purposes, friction is irrelevant for a baseball falling two meters at sea level. But if the distance the ball is dropped is 15 kilometers above the surface of

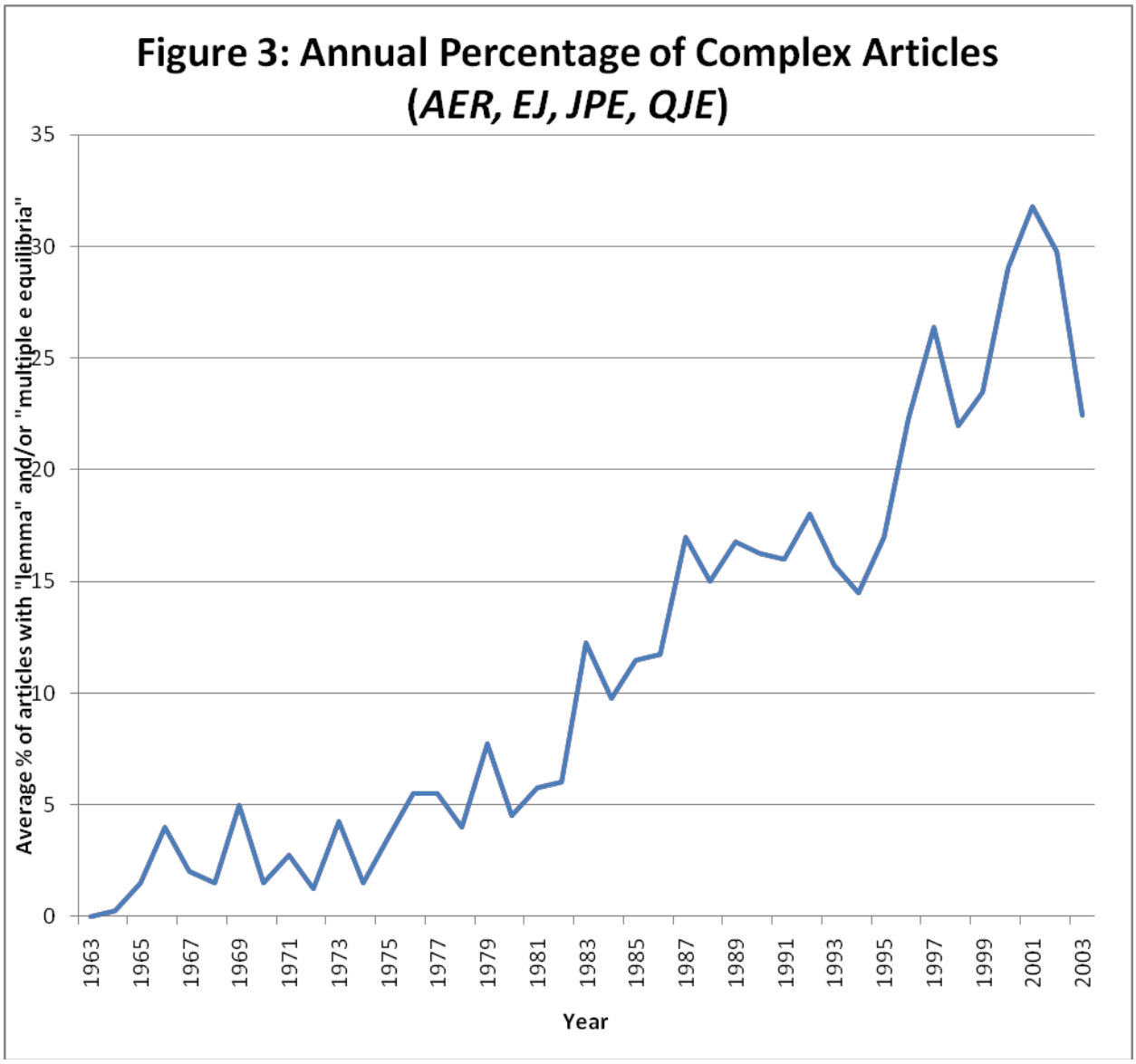
[17]The proverbial butterfly that changes weather in chaos theory is a manifestation of the effects that events occurring in historic ("real") time have over observable phenomena.
[18]It is not impossible to operationalize a general equilibrium model this is constrained, say, to a ***short*** time interval following a ***large*** shock. An excellent example of just such an analysis is the study by Chambers and Gordon (1966) on the impact of the Canadian wheat boom on the Canadian economy during the period from 1901 to1911.

the earth, then friction, wind currents, humidity, and other variables matter; for most purposes experimental falling baseballs data will differ too greatly from that predicted by the model for the model to be useful regardless of its mathematical sophistication.

B. Trends in Complexity and evidence that it inhibits operationalism

In a study of theoretical complexity (2005) we gathered data on the percentages of articles containing the terms "lemma" and/or "multiple equilibrium" at various economic journals.[19] The data showed positive and significant trends in the percentage articles containing these terms from 1963 to 1996 for the: *AER, EJ, JPE, QJE*. Here the series is extended through 2003 and combined the percentages of complex articles for each year for all four journals in an unweighted average. The trend in the mean percentage of articles for each year is provided in Figure 3.

[19] This is unlike our earlier searches that used the term "lemma 1;" the searches for *mathematical complexity* were for the use of the terms "lemma" and/or "multiple equilibrium" in these journals.



The broad measure of mathematical complexity in Figure 3 (it includes both "lemma" and "multiple equilibria") is consistent with the positive time trend of just the word "lemma" illustrated in Figure 1.[20] The question remains: What evidence is there that trends toward mathematical complexity have worked against operationalism; alternatively, what evidence is there for the Gordon hypothesis that complexity and operationalism are negatively related in economic theory?

[20] See Sutter and Pjesky (2007) on "math free" publications in top economics journals.

In another study (2008) we examined the consequences of lemma usage (the proxy for mathematical complexity) from various perspectives. First, we investigated lemma usage in articles in the *AER, EJ, JPE,* and *QJE* that had been cited 500 or more times according to a list compiled by Kim, Morse, and Zingales (2006, p. 15.). Of the 59 articles in these journals that had received 500 or more citations, only one article employed lemmas [by Cho and Kreps (1987), who created two lemmas in their article]. Second, for all the journals on the Kim, Morse, Zingales list of highly cited articles, we compared the relative frequencies of articles containing no lemmas versus those having one or more lemma across two distinct types of analytics: "Economic" versus "Statistical/Econometric". We found a significant difference: While 52% of highly cited articles developing statistical/ econometric analytics contained lemma(s), only 11% of the highly cited journal articles in economics (not econometrics) had lemma(s). These findings dovetail with the Gordon hypothesis that operationalism and mathematical complexity are negatively related in economic analysis (not statistical/econometric analysis); as we explained (p. 346): ". . . this makes intuitive sense: widely cited statistical/econometric analytics generally supply directly or contribute indirectly to econometric tests and techniques for the manipulation of data. These [econometric/statistical] articles are widely cited because what they supply is useful for examining data in articles that are operationalizing theories." The Gordon hypothesis applies to theories that deal with economic phenomena; it does not apply to the pure mathematics whose complexities are essential in the creation of valid statistical/econometric tests. (The proof of the internal consistency of a statistical test is crucial to their utility.)

Our 2005 article uses publications appearing in the *AER* to provide more direct tests of the Gordon hypothesis (again that mathematical complexity and operationalism are negatively related in economic theory). First, we examined the contents of *AER* publications containing the terms "lemma" and/or "multiple equilibria" to the contents of a random sample of *AER* publications, and found that: "The presence of 'lemma' and/or 'multiple equilibria' in an article has a [significant] negative impact on the probability that the article has any empirical content." Secondly, we compared the contents of citations to mathematically complex *AER* publications

(again, proxied by the presence of "lemma" and/or "multiple equilibria") to the contents of citations to the random sample. This allow the assessment of whether those citing the more mathematically complex publications were the same as the random sample. ". . . [T]he presence of the term "lemma" and/or "multiple equilibria" in the source article had a [significant] negative impact on the probability of a citation containing any empirical analysis." (p. 564)

III. Why Mathematically Complexity is Thriving

A. Stylized Facts: Fabricating Reality

The history of economic ideas attests to a cavalier indifference to empirical reality. Vivid illustrations are provided in three separate articles that dispelled hoary economic fictions: Steven N. S. Cheung's (1973) article, "The Fable of the Bees: An Economic Investigation;" Ronald H. Coase's (1974) article "The Lighthouse in Economics;" and S. J. Liebowitz's and Stephen E. Margolis's, "The Fable of the Keys."[21] Cheung dispelled the notion that the external effects of bee keeping would lead to an underproduction of bee keeping services and an inefficient outcome; he did this by investigating the actual markets for honey, bee keeping, and pollination services. Coase showed that lighthouses' services could be privately provided, and not necessarily provided by government; again, Coase investigated the history of lighthouses and navigation, and made his case. Liebowitz and Margolis argued that the QWERTY keyboard as an illustration of the concepts of technological "lock-in" or path dependency had almost no basis in fact. All these papers appealed to objective reality and history. Their primary objective was methodological: To deter economists from *fabricating* reality to fit their theories.[22] In Cheung's words:

[21] "The Fable of the Keys" is a play on Cheung's earlier work, which, in turn, echoes the title of Mandeville's famous early eighteenth century title: "The Fable of the Bees."
[22]See Posner (1993) for more on Coase's methodological approach.

Thus to assume the state of the world to be as one sees fit is not even to compare the ideal with the actual but, rather, to compare the ideal with a fable. . . . My main criticism, rather, concerns their approach to economic inquiry in failing to investigate the real-world situation and in arriving at policy implications out of sheer imagination. (p. 33)

Similarly Coase wrote:

The question remains: how is it that these great men [John Stuart Mill, Henry Sidgwick, and Paul A. Samuelson] have, in their economic writings, been led to make statements about lighthouses which are misleading as to the facts, whose meaning, if thought about in a concrete fashion, is quite unclear, and which, to the extent that they imply a policy conclusion, are very likely wrong? The explanation is that these references by economists to lighthouses are not the result of their having made a study of lighthouses or having read a detailed study by some other economists. Despite the extensive use of the lighthouse example in the literature, no economist, to my knowledge, has ever made a comprehensive study of lighthouse finance and administration. The lighthouse is simply plucked out of the air to serve as an illustration. The purpose of the lighthouse example is to provide "corroborative detail, intended to give artistic verisimilitude to an otherwise bald and unconvincing narrative." (p. 374-375, footnotes omitted)

Liebowitz and Margolis make methodology their ultimate point:

Finally, it is consistent that in a world in which mistakes are frequent and permanent, "scientific approaches" cannot help but make big improvements to market outcomes. In such a world, there is ample room for enlightened reasoning, personified by university professors, to improve on the consequences of myriad independent decisions. What credence can possibly be given to a keyboard that has nothing to accredit it but the trials of a group of mechanics and its adoption by millions of typists? If we use only sterilized models of markets, or ignore the vitality of the rivalry that confronts institutions, we should not be surprised that the historical interpretations that result are not graced with the truth that Cicero asks of historians.

Facts and history matter; asserting a state of the world that contradicts reality, and then constructing a theory to accommodate the fictional world is bizarre at best. Yet this is the protocol of mathematical economics. False assumptions do not disprove a theory; the assumptions' effects depend upon whether they materially affect the hypotheses' operationalism. But absent operational content, what is the point of presenting complex constructions with bizarre assumptions?[23] The following examples are of assumptions ("stylized facts") that

[23] The "lock-in" effects (path dependency) of QWERTY that Paul David espoused in his *American Economic Review* (1985) article was cited in the government's (successful) anti-trust case against Microsoft. Again, there is no creditable evidence of QWERTY's superiority, but it

contradict reality: 1) children are desired because they support their parents in old age (Ehrlich and Lui, 1991); 2) demand curves are upward sloping (Pesendorfer, 1995; Gary S. Becker, 1991[24]); and 3) married men and women uniformly regard the birth of a boy as valuable, while attaching no value to the birth of a girl (Cole, Mailath, and Postelwaite, 1992). All these theories are supported by "stylized facts" and proven to be mathematically consistent.[25] As long as editors and authors persist in fabricating reality, unconstrained by appeals to reproducible and independent information, the fictions persist. They are like the vampire unable to stand the light of day, yet still wreaking havoc with reality.

B. Editorial Animus Towards Critical Commentary

These observations led to concerns about editorial policies that may connect with the persistence of errors, overstatements, or hypotheses based on what someone once said. An

---

lives on as an example of technological path dependency. So a cynical defense of these spurious examples could be that they work in deluding courts, historians, reporters, and the public.

[24] See Gisser, Okten, McClure, and Santoni (2009) on the untenable implications attending Becker hypothesis.

[25] S. Landsburg (1995) challenged the mathematical proof of the existence of multiple equilibria in the growth model of Cole, Mailath, and Postelwaite (1992). Cole, Mailath, and Postelwaite (1995) admitted an error and modified their (1992) assumptions in order to establish valid proofs of the existence of the multiple equilibria. The tenor over the existence of multiple equilibria is represented by the summarizing remarks of Cole, Mailath, and Postelwaite's reply (1995, p. 443):

> In summary: (1) Wealth-is-status equilibria exist for all values of γ, the coefficient of risk aversion. (2) If the initial wealth distribution has $k_0(0)>0$, with a zero measure set of males having initial distribution $k_0(0)$, aristocratic equilibria do not exist for any specification of the utility function of the wife's quality v(j) that has $v'(0)>0$. (3) If the initial wealth distribution has $k_0(0)=0$, aristocratic equilibria may exist. A sufficient condition for the case $\gamma \geq 1$ is Landsburg's condition 1. (4) If the initial wealth distribution has $k_0(0)=0$ and $\gamma=1$, then aristocratic equilibria also exist when condition 1 is violated.

Understandably operationalism is no part of this debate. How can one operationalize: 1) the coefficient of risk aversion γ is it greater, equal to, or less than one; 2) "a zero measure set of males;" 3) the "wife's quality;" and 4) "condition 1"? Beyond these specifics, there is the broader issue that any theory that predicts *multiple equilibria* has inherent difficulties in being tested.

investigation of *The American Economic Review* was revealing; Orley Ashenfelter (editor, 1989) stated that:

> Although the number of articles has now stabilized at about its 1984 level, our publication of notes, comments and replies has decreased steadily since 1985. Both I and my co-editors believe this is a desirable editorial change. Our goal is to increase the number of major, important research papers in the *Review*, and we expect this to come mainly at the expense of our publication of brief notes and comments. (pp. 405-406)

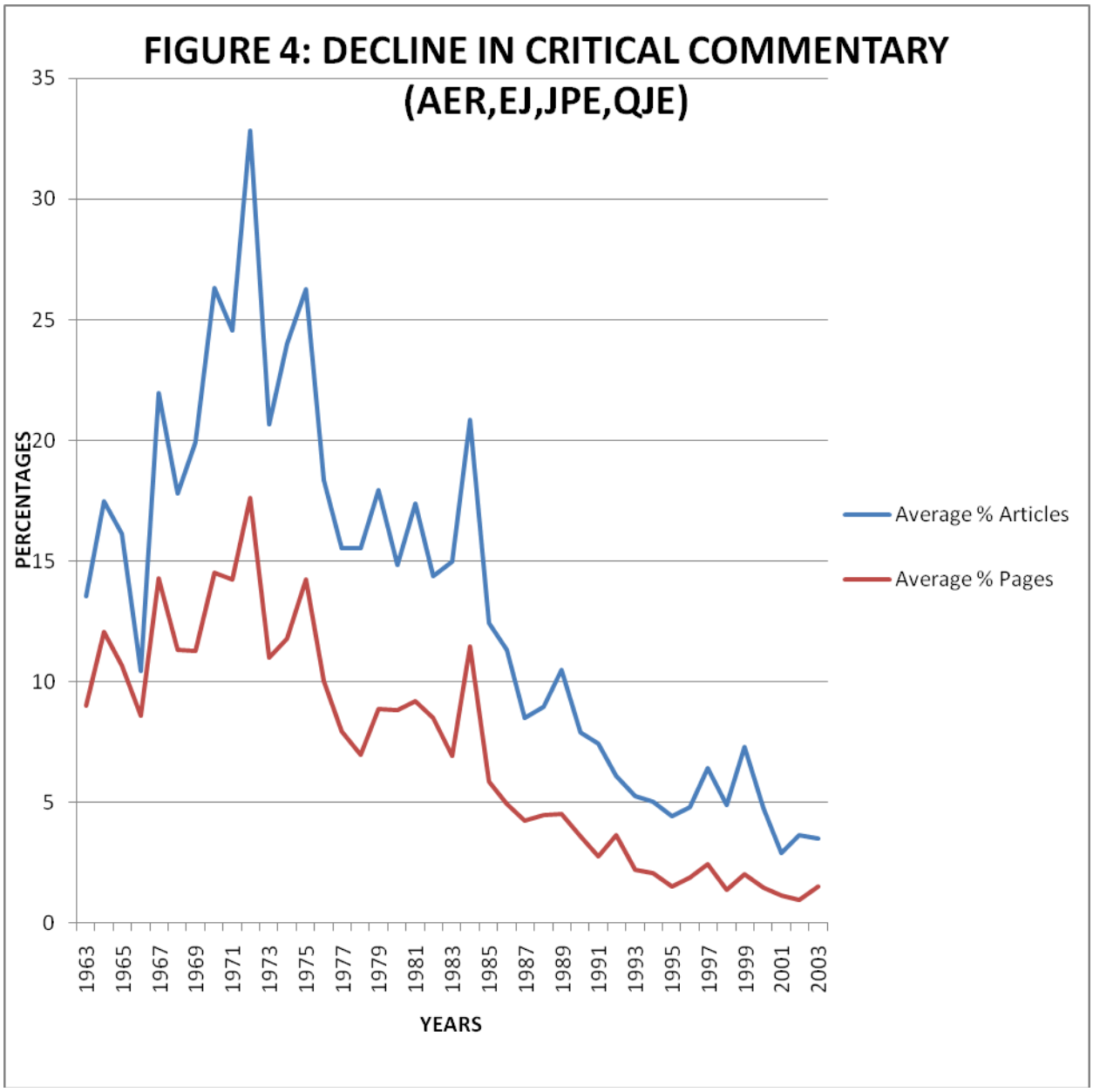
A co-editor of the *Review,* R. Preston McAfee (1996) amplified Ashenfelter's statement: "...the *Review* is **intentionally hostile** to comments and notes, for overall readership for comments tends to be restricted to the readers of the original article." (emphasis added) The editorial policies of the *AER* have changed into ones that value "original" articles highly, and papers that find published papers flawed and/or fallacious worthless. The curtailment of exchange and debate in the form of notes and comments is a hindrance to dispelling error and/or ignorance.[26] If journals accept "stylized facts" as foundations for complex theories and where there is an overt aversion to debate, then we expect to find an increasing tendency to publish non-operational papers that are literally unreal.

C. Evidence of The Decline in Critical Commentary

The decline in critical commentary is not limited to the *American Economic Review*. There is substantial evidence that comments, replies and rejoinders from 1963 to 2003 declined markedly in these major general-interest journals: *AER, EJ, JPE,* and *QJE* (Coelho, De Worken-Eley, and McClure 2005). Figure 4 aggregates the results across these four journals for the

[26] Replication is another way to constrain error; but evidence provided by B.D. McCullough (2007) indicates that archiving practices are crucial and that some editors have been resistant to calls for modifications necessary to insure replicability.

percentage of total articles that are comments, replies, or rejoinders.[27]



Comparing the results depicted in Figures 3 and Figure 4 we see that critical commentary declined dramatically over the same period when mathematical complexity in economic theory was rising. A straightforward explanation of these two opposing trends emerges from our

[27] Over a similar time period, B. Dollery, J. Byrnes, and G. Akimova (2008) found similarly steep declines in percentages of articles and percentages of pages devoted to critical commentary articles in Australian economics journals.

discussion: Given the longstanding tendency of economists to entertain fabricated reality, editorial animus from the 1980s going forward toward critical commentary has opened greater opportunities for the publication of mathematically complex models that fabricate reality. While ours is the only explanation we are aware of that links the trends in Figures 3 and 4, a number of explanations solely about the decline in critical commentary have been advanced.[28]

IV. Concluding Remarks

The evidence supports the Gordon hypothesis: more complex mathematical propositions in economics are less likely to be operational. Critics may argue that we have demonstrated the obvious: Theoretical papers are about theory. This raises two issues: 1) if complex mathematical theories are about economic phenomena, why is there an absence in interest and/or ability to assess them empirically? The Gordon hypothesis provides an answer: When longer chains of mathematical functions are linked together in an economic theory it becomes more likely that the passage of time materially alters the specified relationships in unpredictable directions. We are unaware of any other equally straight forward explanation for the relative deficiency of operationalism in complex mathematics in economics. The second issue raises the more fundamental question: What is the proper domain of economics? Is it Marshall's: "... a study of mankind in the ordinary business of life..." (p. 1), or, alternatively is it a realm where the purity of logic and mathematics is to be pursued unsullied by observational reality? In the "ordinary business of life" we are unlikely to encounter any absolutes be they "Truth," beauty, or mathematical purity. In the tradition of Marshallian economics, the imperfections introduced by

[28] One explanation for the decline may be that economics and the journal literature have reached the apogee of intellectual excellence; reaching excellence explains the increasing emphasis upon minutia as all the major problems in economics have been solved. But recent history suggests that anyone who thinks that all the major economic controversies have been resolved is divorced from reality. Laband, Tollison, and Karahan (2002) suggest this pinnacle of excellence hypothesis as a possible explanation for the decline in commentary; more seriously they offer other explanations: 1) editorial pre-screening may have improved, reducing the need for *ex-post* monitoring by the profession via critical commentary; and 2) editorial rent seeking might be to blame. Alternatively Whaples (2006) suggests is that the decline in critical commentary articles may be due to the dearth of citations that they get relative to regular articles.

measurements, operationalism, experience, and history will temper the use of mathematics. In this tradition theories are conditional upon time, history, and other pertinent circumstances. The search for absolutes is antithetical to conditional statements.

Before the start of the lemma- multiple equilibria era, academic economic journals provided a forum for discussions between economists expressing opposing views on such things as: 1) the "appropriate" balance between the mathematical complexity of hypotheses and their operationalism; 2) the relative merits of "stylized facts" versus observational reality; and 3) academic discourse in the form of comments and notes that addressed empirical verification, the appropriateness of model formulation, and other details that bedevil research. Over the last four decades of the 20th century editorial policies of the major economic journals have expelled the devil of details and have severely restricted debate. Comments, replies and rejoinders in the *AER, EJ, JPE* and *QJE* have gone from peak to trough as editors have become increasingly "hostile" toward perspectives other than the ones they published. If all that is published is correct, this may be a good thing. We are skeptical: *Errare humanum est*. True wisdom recognizes that error is eternal, but then again, we may be wrong.

References

Alchian, A. (1950). "Uncertainty, evolution, and economic theory," Journal of Political Economy 58(3): 211-21.

Alcock, J. (1998). Animal Behavior; 6th edition. Sunderland, Mass.: Sinauer Associates.

Ashenfelter, O. (1989). "Report of the editor," American Economic Review 79(2); 405.

Azariadis, C. (1983). "Employment with asymmetric information," Quarterly Journal of Economics 98 (Supplement): 157-172.

Becker, G. (1991). "A note on restaurant pricing and other examples of social influence on price," Journal of Political Economy 99 (5): 1109-1116.

Boland, L. (1979). "A critique of Friedman's critics," Journal of Economic Literature 17(2): 503-522.

Boland, L. (1984). "Methodology: Reply," American Economic Review 74(4): 795-97.

Buchholz, T. (1990). New Ideas from Dead Economists, New York: Penguin Books.

Chambers, E., and D. Gordon (1966). "Primary products and economic growth: An empirical measure," Journal of Political Economy 74 (4): 315-332.

Cheung, S. (1973). "The fable of the bees: An economic investigation." Journal of Law and Economics XVI(1): 11-33.

Cho, I., and D. Kreps (1987). "Signaling games and stable equilibria," Quarterly Journal of Economics 102 (2): 179-222.

Coase, R. (1974). "The lighthouse in economics," Journal of Law and Economics XVII (October): 357-76.

Coelho, P., and J. McClure. (2008). "The market for lemmas: Evidence that complex models rarely operate in our world," Econ Journal Watch 5(1): 78-90.

Coelho, P., and J. McClure (2006). "Why has critical commentary been curtailed at top economics journals? A reply to Robert Whaples," Econ Journal Watch 3(2): 283-291.

Coelho, P., and J. McClure (2005). "Theory versus application: Does complexity crowd out evidence? Southern Economic Journal 71(3): 556-565.

Coelho, P., D. Klein, and J. McClure (2004). "Fashion cycles in economics," Econ Journal Watch 1(3): 437-454.

Coelho, P., D. Klein, and J. McClure (2005). "Rejoinder to Pesendorfer," Econ Journal Watch 2(1): 32-41.

Coelho, P., J. McClure, and D. Worken-Eley (2005). "Decline in critical commentary in economics: 1963-2004." Econ Journal Watch 2(2): 355-361.

Cole, H., G. Mailath, and A. Postlewaite (1992). "Social norms, saving behavior, and growth," Journal of Political Economy100(6): 1092-1125.

Cole, H., G. Mailath, and A. Postlewaite (1995). "Response to 'aristocratic equilibria,'" Journal of Political Economy 103(2): 439-43.

Cronin, H. (1991). The Ant and the Peacock. New York: Cambridge University Press.

Dollery, B., J. Byrnes, and G. Akimova (2008). "The curtailment of critical commentary in Australian economics," Econ Journal Watch 5(3): 349-51.

Ehrlich, I., and F. Lui (1991). "Intergenerational trade, longevity, and economic growth," Journal of Political Economy 99(5): 1029-1059.

Frazer, W. (1984). "Methodology: Reply," American Economic Review 74(4): 793-94.

Frazer, W., and L. Boland. "An essay on the foundations of Friedman's methodology," American Economic Review 73(1): 129-144.

Girard, A. (2001). "Simultaneous ascending auctions and the Federal Communications Commission Spectrum Auction 35," *The Telecommunications Review*, Noblis Group, pp. 83-91.

Gisser, M., G. Okten, J. McClure, and G. Santoni (2009). "Some anomalies arising from bandwagons that impart upward sloping segments to market demand," Econ Journal Watch 6(1): 21-34.

Gordon, D. (1955; 1968). "Operational propositions in economic theory," In, Readings in Microeconomics, (Edited by William Breit and Harold Hochman). New York: Holt, Rinehart, and Winston, Inc.

Gordon, D. (1955). "Professor Samuelson on operationalism in economic theory," Quarterly Journal of Economics 69(2): 305-310.

Harris, M., and Y. Weiss (1984). "Job matching with finite horizon and risk aversion," Journal of Political Economy 92 (4): 758-779.

Hausman, D. (1989). "Economic methodology in a nutshell," Journal of Economic Perspectives 3(2): 115-127.

Hellwig, M. (1981). "Bankruptcy, limited liability, and the Modiglianai-Miller Theorem," American Economic Review 71 (1): 155-170.

Herberg, H., and M. Kemp (1980). "In defense of some 'paradoxes' of trade theory," American Economic Review 70 (4): 812-814.

Hirsch, A., and N. De Marchi (1984). "Methodology: A comment on Frazer and Boland, I," American Economic Review 74(4): 782-788.

Hirschleifer, J. (1977). "Economics from a biological viewpoint," Journal of Law and Economics XX (1): 1-52.

Hoover, K. (1984). "Methodology: A comment on Frazer and Boland, II," American Economic Review 74(4): 789-92.

Journal Storage, 2000-2002 (*JSTOR*): http://www.jstor.org.

Khan, M., and R. Vohra (1987). "An extension of the second welfare theorem to economics with nonconvexities and public goods," Quarterly Journal of Economics 102 (2): 223-241.

Kim, E., A. Morse, and L. Zingales (2006). "What has mattered to economics since 1970?" NBER Working Paper #12526, National Bureau of Economic Research: Cambridge, MA.

Klein, D., and P. Romero (2007). "Model building versus theorizing: The paucity of theory in the *Journal of Economic Theory*," Econ Journal Watch 4(2): 241-71.

Kuhn, T. (1962). The Structure of Scientific Revolutions. University of Chicago Press.

Laband, D., R. Tollison, and G. Karahan (2002). "Quality control in economics", Kyklos 55(3): 315-34.

Landsburg, S. (1995). "Aristocratic equilibria," Journal of Political Economy 103(2): 434-38.

Laffont, J., and J.Tirole (1986). "Using cost observations to regulate firms," Journal of Political Economy 94 (3): 614-641.

Leamer, E. (1978). Specification Searches: Ad hoc Inferences with Nonexperimental Data. New York: Wiley.

Leijonhuvud, A. (1973). "Life among the econ," Western Economic Review 11(3): 327-337.

Liebowitz, S., and S. Margolis (1990). "The fable of the keys," Journal of Law &Economics XXXIII (April): 1-25.

Liebowitz, S., and S. Margolis (1995). "Path dependence, lock-in, and history," Journal of Law, Economics, & Organization 11(1): 205-226.

Macleod, C. (2002). Book Review of S. Liebowitz's and S. Margolis's: Winners, Losers and Microsoft: Competition and Antitrust in High Technology; in Economic History Society: 384-85.

Malcomson, J. (1984). "Work incentives, hierarchy, and internal labor markets," Journal of Political Economy 92 (3): 486-507.

Marshall, A. (1964). Principles of Economics (8th Ed.). London: MacMillian & Co.

McAfee, R. (1996): American Economic Review editorial correspondence with P. Coelho.

McAfee, R. and J. McMillan (1996). "Analyzing the airwaves auction," Journal of Economic Perspectives 10 (1): 159-75.

McCloskey, D., (1983). "The rhetoric of economics," Journal of Economic Literature 21(2): 481-517.

McCloskey, D. (1994). Knowledge and Persuasion in Economics, Cambridge.

McCullough, B. (2007). "Got replicability? The Journal of Money, Banking and Credit and archive," Econ Journal Watch 4(3): 326-37.

Nagel, E.(1963) "Assumptions in economic theory," American Economic Review 53(2), Papers and Proceedings of the Seventy-Fifth Annual Meeting of the American Economic Association: 211-219.

Pesendorfer, W. (1995). "Design innovations and fashion cycles," American Economic Review 85(4): 771-92.

Pesendorfer, W. (2004). "Response to 'fashion cycles in economics,'" Econ Journal Watch 1(3): 455-464.

Pesendorfer, W. (2005). "Second reply to Coelho, Klein, and McClure," Econ Journal Watch 2(1): 42-46.

Piron, R. (1962). "On 'the methodology of positive economics': Comment," Quarterly Journal of Economics 76 (4): 664-666.

Popper, K. (1934/1959). Logic of Scientific Discovery. New York: Science Editions.

Posner, R. (1993). "Nobel Laureate: Ronald Coase and methodology," Journal of Economic Perspectives 7(4): 195-210.

Roemer, J. (1986). "Equality of resources implies equality of welfare," Quarterly Journal of Economics 101 (4): 751-784.

Ross, S. (1987). "Arbitrage and martingales with taxation," Journal of Political Economy 95 (2): 371-393.

Rotwein, E. (1959). "On 'the methodology of positive economics,'" Quarterly Journal of Economics 73(4): 554-575.

Rotwein, E. (1962)."On 'the methodology of positive economics': Reply," Quarterly Journal of Economics 76 (4): 666-668.

Samuelson, P. (1952). "Economic theory and mathematics–An appraisal," American Economic Review 42(2), Papers and Proceedings of the Sixty-fourth Annual Meeting of the American Economic Association: 56-66.

Samuelson, P. (1965). Foundations in Economic Analysis. New York: Atheneum.

Social Science Citation Index (various dates), Philadelphia, PA: Thompson Scientific.

Spence, M. (1973). "Job market signaling," Quarterly Journal of Economics, 87: 355-74.

Sutter, D., and R. Pjesky (2007). "Where would Adam Smith publish today? The near absence of math-free research in top journals," Econ Journal Watch 4(2): 230-40.

United States Department of Commerce/Bureau of Census (1975). Historical Abstracts of the United States, 93rd Congress, 1st Session, House Document No. 93-78 (Part 2).

Whaples, R. (2006). "The cost of critical commentary in economic journals," Econ Journal Watch 3(2): 275-282.