SNP Best-set Typesetter Ltd.	
Journal Code: AJES	Proofreader: Emily
Article No: 514	Delivery date: 12 February 2007
Page Extent: 44	

What Do Economists Talk About?

A Linguistic Analysis of Published Writing in Economic Journals

By Nils Goldschmidt and Benedikt Szmrecsanyi*

Abstract. The present study aims to add to our knowledge about economic rhetoric by conducting a data-driven analysis of economic academic discourse, both synchronically in its contemporary form, and diachronically over the past four decades. We find (1) that linguistically, economics is clearly an academic genre of its own, (2) that there are at the same time clear differences in vocabulary and style usage across economic journals, and (3) that there have been major developments in economic prose during the past four decades. We argue that there is some, albeit tentative, evidence that the discipline may face an increasing methodological gap.

I

Introduction

Authors such as D. McCloskey (e.g., McCloskey 1983, 1985b), Arjo Klamer (e.g., Klamer 1984), and Uskali Mäki (e.g., Mäki 1988), among others, have sparked a broad debate on the relationship between economics and language.[1] While this discourse is still ongoing, it seems to have arrived at a dead end, or at least turned into a circle: Instead of new analyses, there are a growing number of essays that reflect on the intra-rhetoric approach debate (e.g., McCloskey 1995, 1998; Mäki 1995, 2000), that deliver anecdotal episodes (e.g.,

*Nils Goldschmidt is a Research Associate at the Walter Eucken Institut, Goethestrasse 10, D-79100 Freiburg, Germany; e-mail: goldschmidt@walter-eucken-institut.de. His research interests include history of economic thought, economic methodology and constitutional economics. Benedikt Szmrecsanyi is an Assistant Professor at the English Department of the University of Freiburg, Rempartstrasse 15, D-79098 Freiburg, Germany; e-mail: benedikt.szmrecsanyi@uni-freiburg.de. His main research interest lies with determinants of linguistic variation.

American Journal of Economics and Sociology, Vol. 66, No. 2 (April, 2007).

© 2007 American Journal of Economics and Sociology, Inc.

McCloskey 2001), or that come up with (autobiographical) stories on the origins of the movement (e.g., Klamer 2001)—perhaps the first signs of a decline in research activity. Besides this tradition, it is worth mentioning that there is a second field in which the study of language and the study of economics meet: the relevance of economic principles to the shape and development of human language, viz. the economic approach to language. There is a long and venerable tradition in linguistics that deals with "economy in language."[2] For instance, Stephen Levinson succinctly stated that the reason why so much in human language goes by implicature is that "inference is cheap, articulation expensive, and thus the design requirements are for a system that maximizes inference" (Levinson 2000: 29). Recently, Ariel Rubinstein (Rubinstein 2000), an economist, restated the importance of economy for language a more formalized fashion and pointed out its relevance for understanding how judgments and decisions are formed in economic theory and policy.

In the present study, we propose another approach to economics and language, one that is thoroughly empirical and uses up-to-date linguistic methods.[3] We will conduct (1) a synchronic analysis of contemporary economic scholarly writing and (2) a diachronic analysis of how academic prose has developed and changed since 1965. More specifically, we will examine linguistic variation across published writing, using a sizable database, including several economic journals with differing emphases. In addition, we will use samples of academic prose from three other disciplines (sociology, microscopy, and mathematics) as benchmarks to further specify the "character" of economic prose and to explore how economic journals differ in their self-conception. The research questions that will guide the present study, then, can be summarized as follows:

1. What are some important characteristics of economic scholarly writing? Are there differences within the discipline, that is, between different economic journals?
2. How does terminology usage and the choice of stylistic features in economic scholarly writing compare to usage in other disciplines such as sociology, microscopy, or mathematics? In other words, do economists consider their discipline a rather

socio-scientific discipline, an applied natural science, or a formal-abstractive discipline?
3. Has published economic scholarly writing changed linguistically during the past 40 years? If so, how?

Note, though, that our primary intent is to portray the potential of our method; more research will be necessary to further elucidate the nature of language usage in economic academic discourse. Yet, in spite of the somewhat programmatic nature of this study, we are able to present clear linguistic evidence of how economic academic discourse is different linguistically from other disciplines and how there are divergent methodological approaches in economics, as far as can be inferred from language usage.

II

Data

TO ANSWER THE RESEARCH QUESTIONS laid out above, we composed a corpus (a collection of linguistic data), tailored to make possible a comparison between economics and other disciplines on the one hand and an analysis of differing emphases in economics—now and over time—on the other hand. Our corpus spans seven journals (four of which are economic) and, for economics, the period from the mid-1960s until today, with samples taken from 1965, 1980, and the 1990s. The following data were included in our corpus to represent contemporary economic scholarly writing:

- The *American Economic Review* (henceforth: AER). A leading, top-ranked economic journal in the Social Citation Index, AER is one of the flagship publications of the discipline and prides itself of its high scientific standards and its focus on highly original research. See http://www.aeaweb.org/aer/ for the journal's web site. We included Vols. 80–89 (1990–1999) Issue 3 each (sample size: 1,937,051 words) as well as Vol. 55 Issue 3 (1965; sample size: 165,761 words) and Vol. 70 Issue 3 (1980; sample size: 124,689 words).
- The *Economic Journal* (henceforth: EJ). Dedicated to "promoting the advancement of economic knowledge"

(http://www.jstor.org/journals/00130133.html), EJ "is a key source for professional economists in higher education, business, government service and the financial sector, and represents unbeatable value for those who want to keep abreast in current thinking in economics" (http://www.blackwellpublishing.com/journal.asp?ref=0013-0133). EJ does not publish original research only but also pieces intended for a wider audience and devoted to the reflection of newer, theoretical developments in economics. We sampled Vol. 108 (1998) Issues 446–451; Vol. 109 (1999) Issues 452–459; and Vol. 110 (2000) Issues 460–465 (sample size: about 2,424,655 words) as well as Vol. 75 Issues 298 and 299 (1965; sample size: 169,515 words) and Vol. 90 Issues 358 and 359 (1980; sample size: 277,270 words).

- The *Journal of Economic Perspectives* (henceforth: JoEP) "attempts to fill a gap between the general interest press and most other academic economics journals," the publisher asserts (see http://www.aeaweb.org/jep/). JoEP does not feature original research but, for the main part, summarizes and comments on earlier work. Included in our sample are Vols. 4–13 (1990–1999) Issues 2 and 3 each. This sample has a size of about 1,898,835 words.
- The *American Journal of Economics and Sociology* (henceforth: AJES). This journal "sets no ideological standards for its collaborators or contributors" (Moss 2002: 1): "Today, the exciting and unrelenting encounters between sociology and economics are a natural subject to explore, and AJES continues to publish carefully crafted essays in the social sciences that represent interesting analysis with an eye toward assisting both practitioners and academicians to better understand the world in which we live" (http://www.blackwellpublishing.com/journal.asp?ref=0002-9246). We chose to include AJES in our analysis since it is, we think, at the forefront when it comes to accommodating sociological issues in economics. We included Vols. 49–60 (1990–2001) Issues 2, 3, and 4 each (sample size: about 2,106,384 words) as well as Vol. 24 Issues 2, 3, and 4 (1965; sample size: 109,802 words) and Vol. 39 Issues 2, 3, and 4 (1980; sample size: 99,827 words).

In addition, the following samples, drawn from leading journals of other disciplines—sociology, microscopy, and mathematics—were included for benchmarking purposes. As such, these data will serve to locate economic prose within a threefold continuum of socio-scientific style (sociology), applied natural science style (microscopy), and formal-abstractive style (mathematics).

- The *American Journal of Sociology* (henceforth: AJS). AJS was established in 1885 as the first U.S. journal in its field and is the "leading voice for analysis and research in the social sciences, presenting work on the theory, methods, practice, and history of sociology," the publisher asserts (see http://www.journals.uchicago.edu/AJS/brief.html). We sampled Volume 106 Issue 3 (2000; size: about 130,000 words).
- The *Journal of Microscopy* (henceforth: JOM). In our study, we will take JOM to be a typical representative of writing in the applied natural sciences. It covers "all aspects of microscopy and high-energy *in situ* beam analysis" as well as "the application of microscopical techniques or specimen preparation procedures in an investigation" (see http://www.blackwellpublishing.com/journals/jmi/). We included Vol. 206 Issues 1, 2, 3 (2000) and Vol. 207 Issue 1 (2000). This sample has a size of about 183,000 words.
- The *Journal of the American Mathematical Society* (henceforth: JAMS). JAMS is a renowned mathematical magazine "devoted to research articles of the highest quality in all areas of pure and applied mathematics" (see http://www.ams.org/jams/aboutjams.html). Our sample is made up of Vol. 9 Number 4 (1996) and Vol. 15 Number 4 (2002). Its size is about 290,000 words.

As can be seen, the bulk of our data is from the 1990s, which we treated as a monolithic block.[4] An attempt was made to roughly balance sample sizes across the journals, and to represent each year in the 1990s evenly. For EJ, the latter proved difficult due to document format problems, which is why for this journal only the years 1998–2000 were sampled. The samples we took from the 1965 and 1980 volumes are comparatively smaller than our 1990s samples. A

somewhat pragmatic reason for this is that it is exponentially more difficult to obtain digitized text data of writing published prior to 1990.

Also note that we included *all* textual material in each journal issue, including book reviews, editorial notes, research articles, and so forth.[5] Analysis was performed by linguistic text retrieval software (*WordSmith Tools* Version 3.0). In all, with over 9 million words, our corpus can claim to be quite representative of the scholarly writing subject to analysis in the present study; even our pre-1990 data subcorpus is quite sizable (about 940,000 words). For purposes of comparison, the other two studies that are methodically similar to the present study (Chevalier and Hudson 2001; Conrad 1996) relied on much smaller corpora: Chevalier and Hudson's corpus had a size of 250,502 words, Conrad's analysis is based on a corpus of 48,000 words.

III

Method

CORPUS-BASED STUDIES are empirical analyses of how language is used. In this paper, we will use a standard method in linguistics (parallel studies of other academic genres have been cited in Note 3). In general terms, our method is a so-called variationist discourse-analytical approach. It is methodically similar to multidimensional analyses of register variation, an approach developed by Douglas Biber and his coworkers.[6] Using this method, we will quantitatively investigate academic prose from both a lexical and a stylistic perspective. To this end, we will (1) look at general characteristics of the vocabulary used across journals; (2) investigate into the use of some more specific terminology (listed in Table 1 below); and (3) analyze variation across a number of stylistic features (see Table 2 below).

The tokens in Table 1 are lexical topic markers: They reflect *what* is being written about.[7] To be able to arrive at satisfactory generalizations, we felt it to be necessary that they be also categorized into larger groups or discourse types, of which these tokens are characteristic. To avoid a priori assumptions about this classification, we conducted a factor analysis (Principal Components, Varimax Rotation) of text

Table 1

Topic Markers Studied

Discourse Type	Item	Example
Abstractive-mathematical	ASSUME/ASSUMPTION	"We *assume* risk neutrality and independence of relevant factors." (AER Vol. 80, 1990)
	PROOF	"[. . .] no *proof* has been presented that this hypothesis is the correct one [. . .]" (AJES Vol. 80, 1965)
	THEOREM	"[. . .] the above result is stated as a lemma instead of as a *theorem*." (AER Vol. 89, 1999)
Socio-scientific	CULTURE/CULTURAL	"Byelorussians were and remain *culturally* Russian" (AJES Vol. 60, 2001)
	HISTORY/HISTORICAL	"This period in the *history* of economic thought stands out prominently [. . .]" (AJES Vol. 50, 1991)
	INSTITUTION/ INSTITUTIONAL	"Neither did he outline a theory of governance with *institutional* procedures [. . .]" (AJES Vol. 60, 2001)
	JUSTICE	"*Justice* may even drop out of sight when urgency presses upon the government." (AJES Vol. 50, 1991)
	SOCIETY	"[. . .] to protect the kind of *society* we believe to be worth saving for future generations [. . .]" (AJES Vol. 50, 1991)
	THEORY/ THEORETICAL	"[. . .]a good economic *theory* would identify a specific cause for a societal problem [. . .]" (AJES Vol. 50, 1991)
	METHODOLOGY/ METHODOLOGICAL	"*Methodology* is [. . .] the study of the grounds for employing methods." (AJES Vol. 59, 2000)

Table 1 *Continued*

Discourse Type	Item	Example
Empirical	DATA	"Unfortunately, however, there are no *data* on within-country trade." (AER Vol. 89, 1999)
	EVIDENCE	"Is there any *evidence* that economists behave as if they were maximizing equation (I)?" (AER Vol. 70, 1980)
	METHOD	"The [. . .] procedure used, most commonly the simplex *method*, normally is completed by a computer." (AJES Vol. 50, 1991)
	OBSERVATION/ OBSERVE	"[. . .] exogenous technological developments might have amplified some *observed* changes." (AER Vol. 89, 1999)
	EFFECT	"Let us turn, then, to MM's analysis of the *effect* of taxes." (AER Vol. 55, 1965)
	MODEL	"This is proved in two different *models* of uncertainty." (AER Vol. 70, 1980)
	RESULT	"This *result* also holds under certainty." (AER Vol. 70, 1980)
Archetypally Economic	EQUILIBRIUM	"Our strategy is to specify an *equilibrium* model of the auto industry." (AER Vol. 89, 1999)
	RATIONALITY/ RATIONAL	"The neoclassical theory then assumes only one type of [. . .] homo oeconomicus, who is *rational*." (AJES Vol. 60, 2001)

Table 2

Style Features Studied

Function	Item	Example and Hypotheses
Probability markers, authorial comment (see Swales 1990: 136)	expressions of authorial evaluation and/or intrusion (POSSIBLE/CERTAIN/SURPRISING)	"His unfortunate choice of terminology *possibly* served to alienate a considerable body of opinion" (AJES Vol. 60, 2001)
	modals of expectation, advice, and probability (OUGHT/SHOULD)	"*Should* 'ownership' and 'property' be so defined as to refer to a single essential right?" (AJES Vol. 60, 1965)
Explanative style suggesting high intellectual complexity (see Swales 1990: 115)	causal conjunctions (*because*)	"Hence mathematical determinations are not judgments, *because* mathematics is an a priori discipline." (AJES Vol. 50, 1991)
Theorizing and hypothesizing style	conditional conjunctions (*if/provided that/in case*)	"*If* not sterilized, this will lead to a credit crunch [. . .]." (AER Vol. 89, 1999)
Personal style (see Swales 1990)	first-person pronouns (*I/we*)	"[. . .] *we* feel that our assumption represents a good first approximation." (AER Vol 55, 1965)

frequencies of the above tokens in our 1990s samples across all journals subject to analysis in this study.[8] We extracted three major factors, which accounted for 61 percent, 26 percent, and 8 percent of the total variance, respectively. Tokens with high factor loadings on the first factor, or group, are MODEL, ASSUME/ASSUMPTION, OBSERVE/OBSERVATION, METHOD, EVIDENCE, DATA, EQUILIBRIUM, RESULT, and EFFECT; as is evident, this group is best characterized as *empirical* in nature. Group 2 shows a high loading on THEORY/THEORETICAL, METHODOLOGY/METHODOLOGICAL, HISTORY/HISTORICAL, CULTURE/CULTURAL, RATIONALITY/RATIONAL, JUSTICE,[9] SOCIETY, and INSTITUTION/INSTITUTIONAL; with SOCIETY, JUSTICE, and HISTORY/HISTORICAL showing the highest correlations with the factor, this group struck us as markedly *socio-scientific* in nature. Group 3, finally, is most strongly correlated with THEOREM and PROOF, terminology that seemed to us rather *abstractive-mathematical*.

For analytical reasons, we slightly adapted the classification suggested by the above factor analysis for the purposes of this study. For one thing, we chose to group the tokens EQUILIBRIUM and RATIONALITY/RATIONAL into a fourth, separate group or discourse type, which we tentatively termed *archetypally economic*. This is because these two tokens have a strongly economic connotation that we believe should be accommodated in the present study, given its title. Second, we opted to group ASSUME/ASSUMPTION into the abstractive-mathematical (rather than the empirical) group (factor 3) for primarily semantic and connotational reasons. Note that ASSUME/ASSUMPTION has the third-highest loading on factor 3 anyway (although in absolute terms, the token's loading on factor 1 is higher), so that this grouping decision is certainly justifiable.

Thus, on the basis of a statistical analysis and on analytical grounds, individual lexical topic makers will be associated with one of four major discourse types in the remainder of this study:

1. *abstractive-mathematical discourse*, which we will take to be logical-hypothetical reasoning that proceeds through universal deduction and that is unrelated to observation or experience;
2. *socio-scientific discourse*, which we will consider theory-driven, and often normative, discourse whose primary subject is human interaction;

3. *empirical discourse*, which will, in our diction, be strongly data-driven and inductive discourse;
4. *archetypally economic discourse*, which we will regard as discourse that is characteristic of economics, using signal terms such as *rationality* and *equilibrium*.

Note that the above four discourse types are of course neither clear cut nor discrete; often, borders between them are fuzzy, and any given discourse may be empirical and socio-scientific, for instance, at the same time. Nonetheless, we will employ this four-fold classification of topic markers as a helpful tool for cognition in our analysis.

Items in Table 2, then, are style features: They are indicative of *how* authors write about their subjects. We included probability markers and markers of authorial comment (POSSIBLE/CERTAIN/SURPRISING) and the modals OUGHT and SHOULD,[10] causal conjunctions as markers of explanative style, conditional conjunctions as markers of theorizing and hypothesizing style, and the first-person pronouns I and WE as markers of personal style in our analysis. While topic markers, therefore, are key indicators of "aboutness" and of terminology used, style features are indicators of "howness." Note that these five style features are known to be comparatively prone to linguistic variation, which is why they have been included in this study as candidates for fruitful variation research.[11]

All tokens will be investigated by measuring their frequency, that is, by measuring how often a given token turns up in the texts. Frequency will be conceptualized in *frequency per million words* (pmw), which is a means of normalizing frequencies to differing text lengths. Our overall argument will rest on two key assumptions: (1) The samples we use are representative of the language used in the respective journals, and, by inference, of the language used in their discipline and/or their orientation (due to our corpus size, this assumption is unproblematic); and (2), the more often a token or word is used—and thus, the higher its frequency—the more prominently does the topic that the token denotes figure in academic discourse. Although (2) is a quite safe assumption (and adopts linguistic standard practice in analyses of large amounts of data; see, for instance, Rayson et al. 1997), the reliance on vocabulary frequency

lists has limitations: The method is blind, of course, to *how* a token is used and particularly to the stance an author may take toward that token. Yet, there is undoubtedly a positive correlation between the frequency with which an item is used in a given journal and the degree to which the item is important for the scholarly community of which the journal is part of.

Finally, it needs to be spelled out that we will study text samples in an attempt to generalize the results by inference. This approach necessitates making sure that, at a minimum, our results are unlikely to have occurred by chance alone. This is why we will report the statistical probability that our generalizations are wrong throughout this paper.[12] As pointed out by McCloskey (1985a) and McCloskey and Ziliak (1996), statistical significance by itself does not imply substantial or scientific significance. Needless to say, therefore, the statistical tests we will conduct are meant to make possible sound scientific reasoning, not to replace it.

IV

Results

THIS SECTION PRESENTS our findings. First, we will conduct a global analysis of the vocabulary used in the journals investigated in this study. Second, we will examine the frequencies of the topic and style markers defined earlier to analyze how economic scholarly writing differs from academic discourse in sociology, microscopy, and mathematics. Third, we will complement our analysis of contemporary academic prose by an investigation into how economic academic discourse has changed lexically and stylistically in the past four decades.

A. Vocabulary Usage in General

We will start this section by looking at how the subject matters in each of the four economic journals can best be characterized. Key lexical nouns—lexical nouns whose frequency is unusually high in comparison to other texts—provide a good way to do just that. Table 3 gives—for each economic journal individually—those 10 lexical nouns

Table 3

Key Lexical Nouns in Economic Journals

AER 1990s		EJ 1990s		JoEP 1990s		AJES 1990s	
token	chi-sq.	*token*	chi-sq.	*token*	chi-sq.	*token*	chi-sq.
1. equation	1,539	1. pension	672	1. insurance	1,060	1. land	2,133
2. MODEL	1,097	2. unemployment	604	2. liability	688	2. property	679
3. equilibrium	1,013	3. wage	423	3. health	584	3. poverty	631
4. firm	612	4. employment	368	4. program	451	4. water	613
5. divorce	564	5. policy	343	5. labor	428	5. society	604
6. sample	433	6. rate	324	6. species	408	6. rent	554
7. coefficient	422	7. inflation	317	7. airline	370	7. money	531
8. price	421	8. countries	279	8. enterprises	333	8. housing	497
9. consumption	360	9. bank	258	9. care	333	9. abortion	432
10. variable	358	10. exchange	219	10. immigration	319	10. community	429

Note: At df = 1, all chi-sq. values in this table are statistically highly significant ($p < 0.01$).

whose frequency chi-squared score is highest in comparison to their frequencies in a reference corpus including texts from *all* four journals in the relevant period. In other words, the higher any given lexical noun ranks in Table 3, the more the respective journal is "special" with regard to the usage of the respective lexical noun. (Note that grammatical words such as *a* and *the*, proper nouns, adjectives, adverbs, and verbs were excluded from analysis.)

Table 3 presents evidence that as far as overall vocabulary is concerned, the economic journals clearly differ in their emphases—in short, authors tend to talk about different (economic) aspects depending on the journal in which they publish. Note, for instance the very mathematical-formal terminology in AER (*equation, coefficient, variable)*[13] as opposed to the rather "soft" terminology in AJES, which appears to take issue with rather normative concepts (e.g., *poverty*) or socio-scientific aspects (*land, property, society, community*). This, of course, is indicative of AJES's interest in articles that deal with the American "progressive era" ideas and debates such as land reform, taxation, and the legacy of Henry George. EJ and JoEP, by contrast, seem to focus on descriptive aspects (*bank, insurance, health*) and applied economic policy issues (*policy, (un-)employment, labor*). Yet, it should be pointed out that on the whole, economics *is* a coherent discipline, judging from these profiles. This becomes clear when computing key lexical nouns in our noneconomic benchmark journals. Key words in AJS, the sociology journal, are *friendship*, *neighborhood*, *classroom*, *student, crime, school, race, segregation, resistance,* and *semester*; microscopists publishing in JOMS focus on *cell(s), electron, image, pollen, fluorescence, tissue, specimen, beam, plant,* and *resolution*; key words in JAMS, finally, are *gamma, lemma, lambda, theorem, proof, omega, delta, space, theta*, and *morphism*. Hence, while there are notable differences in vocabulary between economic journals, they do nonetheless form a quite homogenous group when compared to noneconomic journals. Economists—to make a long story short—do not write about *friendship*, *cells*, or the Greek alphabet. Their subject matter is recognizably economic.

Let us also talk briefiy about lexical density across the disciplines. Linguists typically measure lexical density in "type-token rations" (TTRs). TTRs are an indicator for any given text's richness or breadth

Table 4

Standardized TTRs Across Journals

AER 90s	EJ 90s	JoEP 90s	AJES 90s
35.9	39.2	42.2	42.0

of vocabulary. We will assume here that the higher TTRs are, the more diverse the topics that are being written about are.[14] Standardized TTRs for the five journals can be seen from Table 4 below.

JoEP and AJES clearly outperform EJ and, especially, AER in the richness of the vocabulary. This means, by inference, that a broader range of topics is covered and, presumably, more complex structures of argumentation are used in AJES and JoEP than in EJ and particularly AER.[15] Mention should also be made that prose in AJS scores a standardized TTR of 41.6 percent: There is, then, really no substantial difference in breadth of topics covered between AJS and JoEP/AJES. Writing published in our JOM sample scores a TTR of 39.2 percent, while our JAMS sample yields a strikingly low TTR of 25.3 percent (which is of course due to the fact that a highly formalized discipline such as mathematics relies on symbols rather than on lexical content nouns). Judging from our benchmark journals, the more mathematical a discipline, the lower are the TTRs; conversely, the more sociological is any given discipline, the higher the TTRs since socio-scientific arguments, economic policy issues, and explanative description all seem to require greater richness in vocabulary. Of the economic journals in our data, AER is the one with the closest affinity to mathematical writing with regard to lexical density.

B. Some Characteristics of Contemporary Economic Academic Discourse

In this section, we will undertake an analysis of our data with regard to those topic and style markers that were defined in Table 1 and Table 2, respectively. Table 5 gives the frequencies of individual topic markers across journals.

Let us begin by looking at those tokens that we took to be characteristic of abstractive discourse, viz. ASSUME/ASSUMPTION, PROOF,

Table 5

Text Frequencies of Topic Markers (in pmw)

	AER 1990s	EJ 1990s	JoEP 1990s	AJES 1990s	AJS	JOM	JAMS
ASSUME, ASSUMPTION	1,139	839	419	391	217	316	1,556
PROOF	274	127	23	23	—	49	2,289
THEOREM	283	74	42	26	—	114	2,771
Sum Abstractive Markers	**1,696**	**1,040**	**484**	**440**	**217**	**479**	**6,616**
CULTURE, CULTURAL	39	85	111	404	626	191	—
HISTORY, HISTORICAL	202	383	541	707	642	—	24
INSTITUTION, INSTITUTIONAL	227	645	777	920	503	—	—
JUSTICE	15	38	48	229	15	—	—
SOCIETY	103	253	288	782	758	98	—
THEORY, THEORETICAL	1,011	1,613	1,401	1,939	1,747	223	375
METHODOLOGY/METHODOLOGICAL	70	163	118	161	433	82	—

Sum Socio-Scientific Markers	**1,667**	**3,180**	**3,284**	**5,142**	**4,724**	**594**	**509**
DATA	1,632	1,382	1,195	881	1,554	659	28
EVIDENCE	602	627	852	317	356	261	14
METHOD	176	357	210	248	294	2194	217
OBSERVATION/OBSERVE	567	310	204	180	387	496	248
EFFECT	2,165	1,735	1,439	931	2,374	942	3
MODEL	3,597	2,783	1,409	996	3,588	621	372
RESULT	2,122	1,452	951	1,037	1,477	1,269	1,039
Sum Empirical Markers	**1,0861**	**8,646**	**6,260**	**4,590**	**10,030**	**6,442**	**1,921**
EQUILIBRIUM	2,024	1,249	572	214	39	76	—
RATIONALITY, RATIONAL	235	236	260	438	101	—	293
Sum Economic Markers	**2,259**	**1,485**	**832**	**652**	**140**	**76**	**293**

and THEOREM. There are highly significant differences between the seven journals ($df = 6$, $p < 0.01$ for all three tokens) ASSUME/ASSUMPTION, PROOF, and THEOREM are vastly more frequent in JAMS than in any of the other journals. This should surprise no one, given the abstract character of mathematics. In JOM and AJS, these tokens are comparatively infrequent. The four economic journals range in between, but there are clear differences between them ($df = 3$, $p < 0.01$ for all three tokens): On a continuum *more abstractive* > *less abstractive* and, evidently, *more mathematical* > *less mathematical*, there is a hierarchy of AER > EJ > JoEP > AJES such that AER contains the most abstractive terminology among the four economic journals and AJES the least. Writings published in AER, hence, appear to have to most mathematical mode of argumentation.

A similar hierarchy holds for the typically economic token EQUILIBRIUM. More or less absent from the noneconomic journals—economists might be surprised to learn that EQUILIBRIUM is not a mathematical term[16]—economists make frequent use of this token. However, there are significant differences ($df = 3$, $p < 0.01$) in that frequencies are stratified such that AER > EJ > JoEP > AJES. In AER, for instance, EQUILIBRIUM is about nine times as frequent as in AJES. RATIONALITY/RATIONAL, on the other hand, is also absent from the noneconomic journals while it is used pretty evenly across the four economic journals ($df = 6$, $p < 0.01$), with a slight tendency for it to be comparatively frequent in AJES. This could mean that in AJES, authors tend to refiect more than elsewhere on this token, which signifies an often presupposed standard assumption in economics.

The hierarchy AER > EJ > JoEP > AJES also obtains for our "empirial" tokens DATA, EVIDENCE, METHOD, OBSERVE/OBSERVATION, EFFECT, MODEL, and RESULT. The distribution of all of these tokens across journals is highly significant ($df = 6$, $p < 0.01$). As was to be expected with regard to our noneconomic benchmark data, empirical topic markers are relatively frequent in AJS and, albeit slightly less so, in JOM. At the same time, this terminology—implying empirical testability—is evidently rather foreign to mathematical prose. As far as economic scholarly prose is concerned, the differences between the economic journals are significant at $p < 0.01$ ($df = 3$) for each individual token in this group. Discourse in AER is "most empirical" and discourse in AJES "least

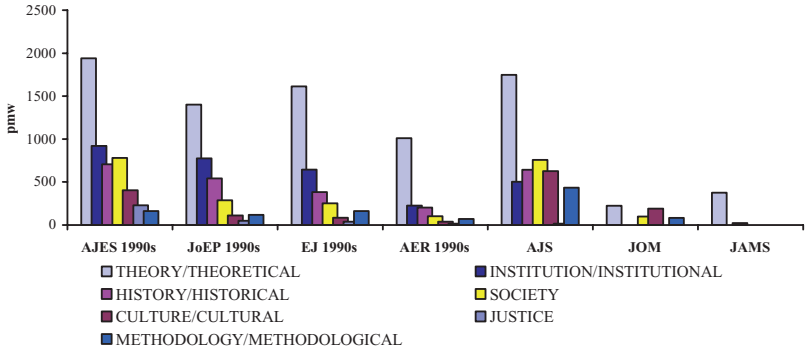
empirical." EJ is in this respect more like AER, while JoEP is more like AJES. With regard to individual tokens in this group, it is noteworthy that MODEL is about equally common in both AER and AJS. The term is used differently in sociology and economics, though: Sociologists use MODEL predominantly to refer to statistical MODELS (for instance, "[T]his article develops a statistical *model* that tests whether [. . .] student background characteristics are more significantly associated with the high incidence of resistant behaviour." AJS Vol 106, 2000), while economists, more often than not at least, use the term to refer to economic models (for instance, "A second example of the use of Theorem 4 is based on a *model* developed by Oliver Hart and John Moore (1990) to study the optimal allocation of ownership rights." AER Vol. 84, 1994) and/or apply standard models of economic theory (e.g., the Heckscher-Ohlin-model) for explaining economic phenomena. Also, it is curious that only in JOM does METHOD play a major role. The much lower frequencies of the token in the economic journals may hint to a weakness of economic reasoning that is supposedly one of its strengths: methods to solve economic policy problems.

We will now move on to terminology that has a social science feel to it: THEORY/THEORETICAL, INSTITUTION/INSTITUTIONAL, HISTORY/HISTORICAL, SOCIETY, CULTURE/CULTURAL, JUSTICE, and METHODOLOGY/METHODOLOGICAL. Its distribution across journals, which is statistically highly significant for each individual token ($df = 6$, $p < 0.01$ across all journals; $df = 3$, $p < 0.01$ across the economic journals), is displayed in Figure 1.

Among the noneconomic journals, AJS (of course) leads JOM and JAMS in the frequency of these items by a wide margin. Among the economic journals, there is a continuum of AJES > JoEP > EJ > AER in that AJES contains the highest share of socio-scientific terminology and AER definitely the lowest. Once again, JoEP appears to be roughly more like AJES and EJ roughly more like AER with regard to the frequencies of most tokens. Mention should also be made that except for CULTURE/CULTURAL and METHODOLOGY/METHODOLOGICAL, there is a tendency for AJES to contain even more of this terminology than AJS (only the differential with regard to INSTITUTION/INSTITUTIONAL, however, reaches statistical significance at $df = 1$, $p < 0.01$). In all, it is worth noting that THEORY/THEORETICAL is relatively frequent in economic prose. This is an indication that referring to theories is important to

Figure 1

Text Frequencies of "Social Science" Topic Markers (in pmw)



economists, more than for authors publishing in JOM and JAMS (disciplines that have somewhat fixed and generally acknowledged standards at their disposal). Tokens such as INSTITUTION/INSTITUTIONAL and HISTORY/HISTORICAL are indicative of social science communities, and SOCIETY is of course primarily a sociological term. Also, CULTURE and JUSTICE play a rather limited role in contemporary economic prose. According to the distribution of the token METHODOLOGY/METHODOLOGICAL, methodological debates do not seem to take place much in contemporary economics, in stark contrast to contemporary sociology.

We will now turn to the usage of style features. Table 6 presents an overview of their frequencies across disciplines and journals.

The frequency distribution of conditional conjunctions (*if/provided that/in case*) across journals is statistically significant ($df = 6$, $p < 0.01$). It is salient from Figure 2 below that frequent usage of these conjunctions are clearly a tell-tale sign of mathematical writing, according to our JAMS sample.

The reason, of course, is that conditional conjunctions are required in highly hypothetical discourse due to their *irrealis* semantics. Mathematicians, then, often need to state conditions for which their theorems hold. Note that there are significant differences between the four economic journals with regard to the frequencies of conditional

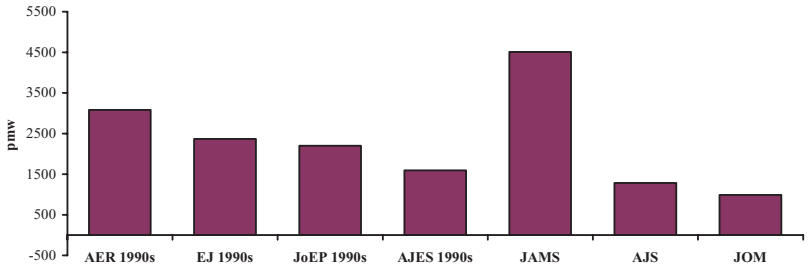
Table 6

Text Frequencies of Style Features (in pmw)

	AER 1990s	EJ 1990s	JoEP 1990s	AJES 1990s	AJS	JOM	JAMS
Expressions of authorial intrusion (POSSIBLE/CERTAIN/SURPRISING)	737	831	891	757	711	936	506
Causal conjunctions (*because*)	967	837	1,176	940	1,083	474	275
Conditional conjunctions (*if/provided that/in case*)	3,084	2,369	2,199	1,595	1,284	898	4,512
First-person pronouns (*I/we*)	3,677	4,909	1,119	2,134	5,266	2,112	3,087
Modals (OUGHT/SHOULD)	585	769	952	879	487	615	186

Figure 2

Text Frequencies of Conditional Conjunctions (in pmw)



conjunctions ($df = 3$, $p < 0.01$). Given the above benchmark, then, we can state that of the four economic journals, AER is the "most mathematical" with regard to usage of conditional conjunctions, and AJES the least. Articles published in EJ and JoEP have roughly equal shares of conditional conjunctions.

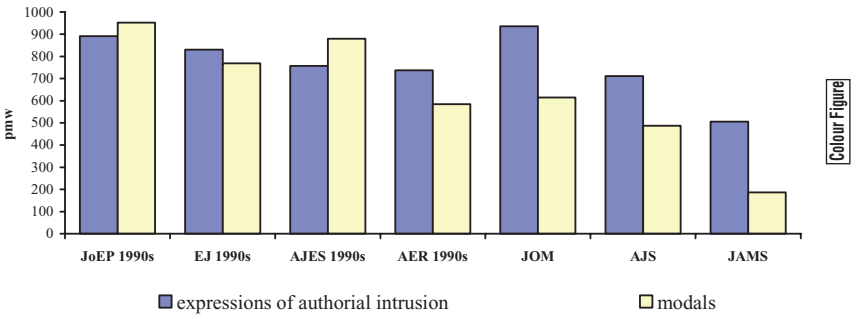
We assumed both expressions of authorial intrusion (POSSIBLE/CERTAIN/SURPRISING) and modals (OUGHT/SHOULD) to be linguistic devices used by authors to make evaluative comments. Usage of these features, again, differs across journals ($df = 6$, $p < 0.01$ for both of them). As can be seen from Figure 3, among the noneconomic journals in our corpus, both of these style features are characteristic of writing in microscopy. By inference, thus, they are characteristic of empirical, descriptive discourse.

It appears, then, that microscopists frequently comment on and express their attitude about their findings. Among our economic journals, authors publishing in JoEP seem to be most ready to use evaluative language. Given JoEP's mission, this is much as one would have expected.

Authors publishing in AER, by contrast, seem to be the least likely to use evaluative language. These differences, again, are statistically highly significant ($df = 3$, $p < 0.01$ for both features). Mention should also be made that in our data, expressions of authorial intrusion and modals are least frequent in mathematical prose (JAMS)—therefore, AER is once again the economic journal with the strongest affinity to mathematical style.

Figure 3

Text Frequencies of Expressions of Authorial Intrusion and Modals (in pmw)



A quite similar relationship among the economic journals holds for causal conjunctions ($df = 6$, $p < 0.01$ across all journals; $df = 3$, $p < 0.01$ among the economic journals) that we assumed to be indicative of a rather explanative style. Among the noneconomic journals, causal conjunctions are most frequent in AJS and least frequent in JAMS; to use a catch phrase, sociologists seem to reason whereas mathematicians state conditions. As far as economics is concerned, causal conjunctions are most frequent in JoEP and they have comparable frequencies in AJES, EJ, and AER. JoEP thus stands out fpr the frequency of devices commonly associated with reasoning style.

The distribution of first-person pronouns, finally, is statistically highly significant as well, both across all journals ($df = 6$, $p < 0.01$) and across the economic journals in our corpus ($df = 3$, $p < 0.01$). What the numbers reveal is that among the noneconomists, sociologists are the ones who employ first-person narration most frequently. First-person narration is rarer in economics, but among economists, the authors in EJ seem to employ it most often by a wide margin. In contrast, first-person narration is markedly rare in JoEP.

To sum up this section: Linguistically, the four economic journals clearly form a genre of their own. This is evident from the similar usage of key lexical nouns as well as from the distinctive usage of RATIONALITY/RATIONAL and EQUILIBRIUM. Yet, these journals apparently do

have different emphases: (1) They cater to different target groups and language communities; (2) the research they publish has different aims; and (3) they encourage specific interpretations of economic facts. They also have different "meta characters": scientific, mathematical, or explanative-descriptive. Now, is economics linguistically rather like sociology, mathematics, or the natural sciences? It is true that some specimens of economic writing—particularly those published in AER—have traits that bear a remarkable affinity to writing published in JAMS, both lexically and stylistically. But our discussion of empirical and socio-scientific topic markers in particular has demonstrated that on the whole, there are a good deal more parallels to sociological academic prose and to sociological modes of argumentation in economic academic discourse. In other words, economic journals may (linguistically) slightly align themselves to other disciplines, but there are severe limits to mimicking the "exact sciences." We believe, too, that our analysis has indicated that the practical application of theoretical insights is a special problem of economics. Unlike in the natural sciences, there is no such thing as "simple" solutions in economic policy matters, in that it is impossible to derive agendas from pure theory. Along these lines, consider the robust, pervasive differences between the economic journals and JOM. Our linguistic analysis, therefore, suggests that, because at heart, economics has to take issue with cultural and social phenomena, it cannot rely entirely on linguistic devices designed to deal with mathematics or the natural sciences—even if the purpose is modeling reality.

C. Economic Scholarly Writing Since 1965: Some Developments

This section will portray language usage in AER, EJ, and AJES—those economic journals for which we collected longitudinal data—over time. To this end, we will draw on data from 1965, 1980, and the 1990s. To make sure that the diachronic developments we observed are really specific to economic academic prose and not due to some general changes in written English or English academic prose specifically (which would be uninteresting in terms of the present study), we checked our findings against two standard linguistic reference corpora, the Lancaster–Oslo/Bergen (LOB) corpus from the early

1960s and its 1990 counterpart, the Freiburg–Lancaster–Oslo/Bergen (F-LOB) corpus.[17] Both corpora contain academic prose subsections of about 200,000 words each. The idea was that if there were any general stylistic or lexical changes in English scholarly writing, these would be reflected in LOB and F-LOB. It turned out that with regard to most of the tokens subject to analysis over time, there were no significant differences between LOB and F-LOB and hence no overall changes in English academic prose. The handful of cases that were significant will be mentioned as we discuss the respective tokens below.

Let us begin by a discussion of some interesting developments with regard to style features, the frequencies of which over time can be seen from Table 7.

We have presented evidence earlier that conditional conjunctions (*if /provided that/in case*) are tell-tale signs of "more mathematical" discourse. According to our data, there is a marked difference between AER on the one hand and AJES and EJ on the other with regard to conditional conjunctions. In AER, their usage peaked in 1980 such that these markers' frequency was lower in the 1990s than in 1980, but still significantly higher in the 1990s than in 1965 ($p < 0.01$ for the periods between 1965 and 1980, 1980 and the 1990s, and the period from between 1965 and the 1990s as a whole). Hence, AER has become significantly "more mathematical" in style over the years, though that development reached its climax in 1980. Conversely, conditional conjunctions decreased significantly in frequency in the period from 1965 until now in both AJES ($p < 0.01$) and EJ ($p < 0.01$). AJES and EJ, therefore, have become significantly "less mathematical" in style since 1965, according to our data.

Second, first-person narration became significantly more frequent over time in AER ($p < 0.01$ for every individual period and the 1965 vs. 1990s period) and significantly less frequent over time in EJ ($p < 0.01$ for every individual period and 1965 vs. 1990s). There were no significant developments with regard to usage of first-person narration in AJES. Hence, while AER became more personal in style, EJ became less so.

Third, modals (OUGHT/SHOULD) and expressions of authorial intrusion (POSSIBLE/CERTAIN/SURPRISING) have become significantly less frequent over time in both AER and EJ ($p < 0.01$ for 1965 vs. 1990s in

Table 7

Style Features: Frequencies Over Time (in pmw)

	AER 1965	AER 1980	AER 1990s	EJ 1965	EJ 1980	EJ 1990s	AJES 1965	AJES 1980	AJES 1990s
Expressions of authorial intrusion (POSSIBLE, CERTAIN, SURPRISING)	1,327	826**	719**	1,327	1,075*	831**	1,002	922	757
Causal conjunctions (*because*)	833	898	967	1,026	700**	837*	829	721	940*
Conditional conjunctions (*if/provided that/in case*)	1,653	3,112**	3,011**	3,191	2,647**	2,369**	2,204	1,412**	1,595
First-person pronouns (*I/we*)	1,762	3,922**	7,656**	8,418	6,849**	4,977**	3,461	3,356	3,672
Modals (OUGHT/SHOULD)	1,207	585**	585	1,215	1,010*	769**	865	912	879

Differences to the preceding period: *significant at the 0.05 level; **significant at the 0.01 level.

both journals); in AJES, there were no substantial changes in frequency. Thus, in the period from 1965 to today, authors publishing in AER and EJ became significantly more reluctant to subjectively comment on their subject matter.

Fourth, the causal conjunction *because* was significantly more frequent in EJ in 1965 than today ($p < 0.01$) and significantly less frequent in AJES in 1965 than today ($p < 0.01$). We could observe no significant usage change of this style feature in AER. Hence, writings published in EJ became "less reasoning" over time, while writings published in AJES became "more reasoning."

Going on to topic markers, Table 8 summarizes their frequencies. Let us begin by discussing abstractive terminology (ASSUME/ASSUMPTION, PROOF, and THEOREM). As a group, abstractive tokens were used significantly less frequently in 1965 than in the 1990s in AER (1965 vs. 1990s: $p < 0.01$), and vice versa for EJ (1965 vs. 1990s). No significant decrease or increase could be observed for AJES. As far as individual tokens in the group are concerned, frequency changes of ASSUME/ASSUMPTION roughly mirror that of the group as a whole in each journal. PROOF steadily became more frequent over time in both AER and EJ (1965 vs. 1990s: $p < 0.01$ for both journals), as did THEOREM (1965 vs. 1990s: $p < 0.05$ for both journals). In sum, EJ seems to have become less abstractive in terminology over time, and AER clearly more abstractive. Note now in particular that THEOREM grew a lot more common in the rather explanative EJ from 1965 to 1980. This means that in this period, a number of economic insights have apparently been summarized into fixed argumentative patterns, maybe for pedagogical reasons. This trend is not observable for the period between 1980 and the 1990s.

Next let us discuss developments in markers of social science discourse. The sum of their frequency scores across journals is displayed in Figure 4.

Developments in this marker group as a whole are generally statistically significant at $p < 0.01$; the only exception is the period from 1980 to the 1990s in AER. To summarize, socio-scientific terminology seems to have temporarily dipped in 1980 in AJES and AER. By contrast, there was a temporary surge in 1980 in EJ, according to our data. When comparing the 1965 AER sample to our contemporary AER

Table 8

Topic Markers: Frequencies Over Time (in pmw)

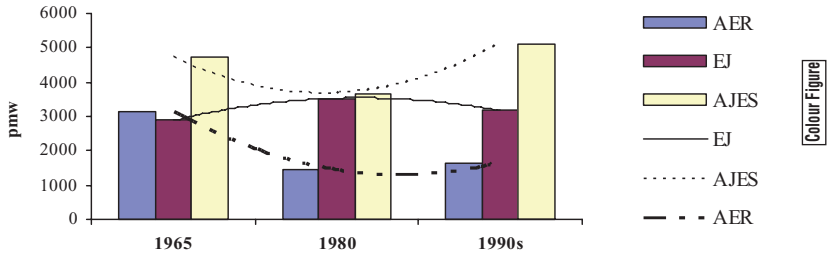
	AER 1965	AER 1980	AER 1990s	EJ 1965	EJ 1980	EJ 1990s	AJES 1965	AJES 1980	AJES 1990s
ASSUME/ASSUMPTION	700	1,500**	1,139**	1,304	815**	839	355	321	391
PROOF	60	104	274**	18	40	127**	27	20	23
THEOREM	211	217	283	24	141**	74**	—	20	26
Sum Abstractive Markers	**971**	**1,821****	**1,696**	**1,346**	**996****	**1,040**	**382**	**361**	**440**
CULTURE/CULTURAL	60	8*	39	53	29	85**	1,047	190**	404**
HISTORY/HISTORICAL	652	104**	202*	519	527	383**	537	411	707**
INSTITUTION/ INSTITUTIONAL	422	112**	227**	260	314	645**	947	321**	920**
JUSTICE	12	—	11	6	32	35	36	160**	196
SOCIETY	205	265	103**	313	191	253	1,002	1,012	782*
THEORY/THEORETICAL	1,719	954**	1,011	1,693	2,308**	1,613**	1,066	1,523**	1,939**
METHODOLOGY/ METHODOLOGICAL	66	8*	70**	41	133**	163	82	60	161*

Sum	**3,136**	**1,451****	**1,663**	**2,885**	**3,534****	**3,177****	**4,717**	**3,677****	**5,109****
Socio-Scientific Markers									
DATA	869	1,668**	1,632	596	559	1,382**	738	511*	881**
EVIDENCE	350	409	602**	313	490**	627**	182	210	317
METHOD	374	112**	176	531	451	357*	355	210*	248
OBSERVATION/OBSERVE	235	401*	567*	112	170	310**	109	110	180
EFFECT	929	1,307**	2,165**	1,009	920	1,735**	492	761*	931
MODEL	1,158	3,040**	3,597**	1,723	2,189**	2,783**	109	861**	996
RESULT	1,104	1,764**	2,122**	908	1,165*	1,452**	647	831	1,037*
Sum Empirical Markers	**5,019**	**8,701****	**10,861****	**5,192**	**5,944****	**8,646****	**2,632**	**3,494****	**4,590****
EQUILIBRIUM	489	1,516**	2,024**	903	848	1,249**	18	170**	214
RATIONALITY/RATIONAL	78	241**	235	41	170**	236*	146	160	438**
Sum Economic Markers	**567**	**1,757****	**2,259****	**944**	**1,018**	**1,485****	**164**	**330***	**652****

Differences to the preceding period: *significant at the 0.05 level; **significant at the 0.01 level.

Figure 4

Text Frequencies of Social Science Topic Markers Over Time (in pmw)



sample, it turns out that socio-scientific terminology was vastly more pervasive in 1965 than it is today (1965 vs. 1990s: $p < 0.01$). Conversely, authors publishing in EJ and AJES use slightly more of that terminology today than they did in 1965 (1965 vs. 1990s: $p < 0.05$ for both journals).[18] Individual tokens in this group are:

1. JUSTICE, which is significantly more frequent now than in 1965 in both AJES (1965 vs.1990s: $p < 0.01$) and EJ (1965 vs. 1990s: $p < 0.05$);
2. METHODOLOGY/METHODOLOGICAL, whose frequency increased significantly from 1965 to today in both AJES (1965 vs. 1990s: $p < 0.05$) and EJ (1965 vs. 1990s: $p < 0.01$);
3. HISTORY/HISTORICAL, which is significantly more frequent now than in 1965 in AJES (1965 vs. 1990s: $p < 0.05$) and significantly less frequent now than in 1965 in both AER (1965 vs. 1990s: $p < 0.01$) and EJ (1965 vs. 1990s: $p < 0.01$);
4. SOCIETY, which decreased significantly in frequency since 1965 in AER (1965 vs. 1990s: $p < 0.01$) and AJES (1965 vs. 1990s: $p < 0.01$). There were no significant changes in EJ;
5. INSTITUTION/INSTITUTIONAL, whose frequency increased in EJ (1965 vs. 1990s: $p < 0.01$) and decreased in AER (1965 vs. 1990s: $p < 0.01$) over the past four decades. In AJES, the token dipped markedly in frequency in 1980 (1965 vs. 1980: $p < 0.01$) before resurging again in the 1990s;
6. CULTURE/CULTURAL underwent no significant changes in both AER and EJ. In AJES, this token's frequency decreased by some 80

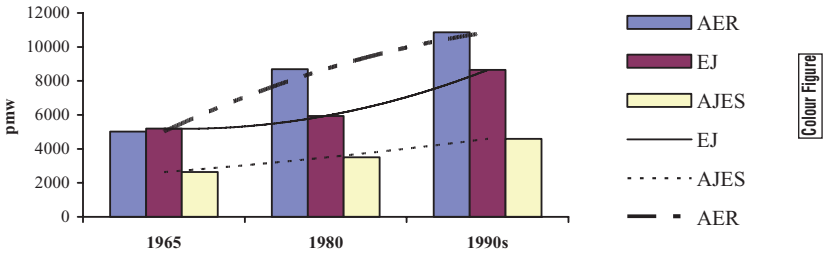
percent from 1965 to 1980 (1965 vs. 1980: $p < 0.01$) while it became more frequent again in the period from 1980 to the 1990s (1980 vs. 1990s: $p < 0.01$). Still, culture/cultural is roughly only half as often used in AJES now than in 1965 (1965 vs. 1990s: $p < 0.01$);

7. THEORY/THEORETICAL, which increased in frequency in AJES (1965 vs. 1990s: $p < 0.01$) and decreased in AER (1965 vs. 1990s: $p < 0.01$) since 1965. In EJ, the token's frequency surged temporarily in 1980 (1965 vs. 1965: $p < 0.01$).

To put it in a nutshell: While socio-scientific terminology is used much more often in AJES than in AER, we can observe in both journals a significantly reduced interest in social science issues in 1980 than in either 1965 or the 1990s. The opposite tendency, which we observed in the EJ texts, emphasizes EJ's reflexive character, in that trends are probably absorbed after some time lag. Three more specific observations merit attention. First, it is worth noting the marked usage decrease over time of CULTURE/CULTURAL, even in AJES. We would conjecture, however, that the uniform frequency increase of the token in the period between 1980 and the 1990s (albeit being one from a very low frequency level to a low frequency level only) will continue in this decade. Second, the decrease of HISTORY/HISTORICAL, particularly in AER but also in EJ, is surprising. It is conceivable that the rather mathematically oriented journals increasingly neglect historical, and hence societal, developments, which is probably also why the history of economic thought is receiving less and less attention in these journals. AJES, conversely, is a journal that is supposed to take a critical stance toward the discipline and thus displays an increasingly socio-scientific orientation, not only with regard to HISTORY/HISTORICAL but also with regard to INSTITUTION/INSTITUTIONAL (which mirrors the growing importance of institutional/constitutional economics). Third, we would like to underline the increasing popularity of the term METHODOLOGY/METHODOLOGICAL, which may mean that methodological considerations are becoming more important in economics. Moving on to empirical terminology, Figure 5 illustrates accumulated frequencies of these markers over time.

Figure 5

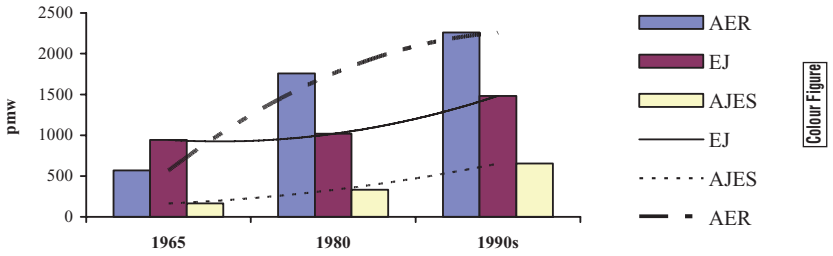
Text Frequencies of Empirical Topic Markers Over Time (in pmw)



In all three journals, there is a very robust trend that empirical tokens on the whole have become more frequent over time (for every individual period and 1965 vs. 1990s: $p < 0.01$). Also, at least for the period from 1965 to the 1990s as a whole, this frequency increase holds for every individual marker except one: METHOD. METHOD steadily decreased in frequency over time and is now significantly less often used than in 1965 in all three journals (1965 vs. 1990s: $p < 0.01$),[19] a development that seems to be complementary to the increasing popularity of METHODOLOGY/METHODOLOGICAL in economic prose. Assuming that the steadily increasing usage of empirical terminology reflects an actually stronger reliance on empirical methods as well, and not only an increasing rhetorical pretension of empirical methods where there are none,[20] economic journals seem to have implemented the often-heard demand for more empirical and econometrical research and education. Along these lines, economists appear to increasingly seek clear solutions: Consider the rise over time of RESULT in our longitudinal data. This contrasts markedly with the weakness of economic journals when it comes to (definite) economic policy proposals (as we have seen in Section IV, B), as compared to application-oriented competence in JOM. Also, note that the more "sociological" an economic journal gets, the less "observations" appear to be made, as the distribution of the token OBSERVE/OBSERVATION in AJES suggests. Conversely, the token is in fact frequent in our sociological journal, AJS.

Figure 6

Text Frequencies of Economic Topic Markers Over Time (in pmw)



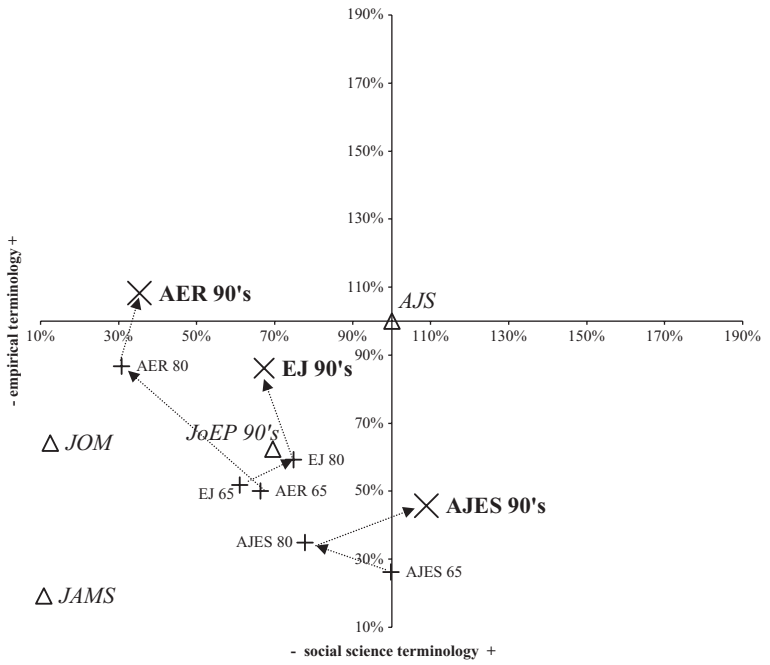
Summarized text frequencies over time of the two typically economic tokens EQUILIBRIUM and RATIONALITY/RATIONAL, finally, are displayed in Figure 6.

As Figure 6 illustrates, text frequencies of EQUILIBRIUM and RATIONALITY/RATIONAL increased steadily over time. For the period from 1965 to the 1990s as a whole, this increase is statistically significant for the two tokens individually in all three journals (1965 vs. 1990s: $p < 0.01$). Across the board, therefore, economists seem to use EQUILIBRIUM and RATIONALITY/RATIONAL a lot more often now than in 1965. This could point to the emergence of an economic core of linguistic—and hence also thematic—unity, consisting of distinctively economic terminology. Two interpretations are possible: There is growing intradisciplinary competence in economics to deal with specifically economic facts; or economists have been severing their ties from social science discourse.

Our findings in this section highlight two tendencies in economic writing: (1) Over time, there has been a growing "economization" of how economists deal with economic phenomena, according to our data. This can be observed in the increasing usage of economic and empirical terminology. AER is clearly leading in this development, and it appears to combine "economization" with formalization. (2) There seems to be a growing methodological gap in economics. Social science dimensions are covered in those journals that have an affinity to this orientation (in our case, AJES). More formally oriented journals—in our case, AER and, with some time lag, EJ—focus

Figure 7

Economic Journals Over Time: Frequency Scores of Social Science Terminology and Empirical Terminology (Percent Deviations from Frequencies in AJS)



increasingly on mathematical argumentation, a point that our discussion of abstractive and style markers has brought home. In this context, we believe that a good way to summarize major changes in economic academic prose since 1965 is Figure 7. It plots the "path" that economic journals have taken since 1965 on two scales (or dimensions), the frequency scores of empirical terminology and the frequency scores of socio-scientific terminology. The reference point in Figure 7 is determined by the frequencies of our contemporary AJS sample scores on these two scales.

Let us take issue with two trends obvious from Figure 7. First, note that AER, EJ, and AJES were demonstrably closer together (at least as

far as a socio-scientific and empirical terminology is concerned) in 1965 than they are now, and that they have been moving in somewhat different directions, even if in all three journals there was a drift toward more empirical terminology. This does not necessarily mean that the discipline is falling apart. It bears evidence, though, that terminology usage has become more and more heterogeneous—there is now a fundamental split between AJES and AER in particular. It seems that economists are increasingly unable to employ the same terminology to deal with identical phenomena, namely, economic reality. Second, it is worth pointing out that by and large, economic scholarly writing is located in the lower left quadrant in Figure 7. This means that economic scholarly prose tends to be both less empirical and less "sociological" or social science–like terminologically than the reference journal AJS. In sum, AER combines high "empiricalness" with a low score on the social science scale, AJES combines a high score on the social science scale with low "empiricalness," and EJ and JoEP occupy middle positions.

V

Conclusion

WE HOPE TO HAVE DEMONSTRATED four major points in this study:

1. When economics is mutually understood by authors and readers as a mathematical or applied science, affinities to more mathematical or applied genres are observable (as we observed for AER). When readers expect authors to reflect on the history of economic thought and to take part in societal discourse, affinities to more sociological disciplines are observable, as we have shown for AJES. EJ and JoEP, finally, seem to be intended for a readership demanding rather explanative and/or descriptive discussion.
2. Authors, in academic communities, are not free to use expressive means of their choosing—rather, they are pressured to imitate (a) the style and mode of argumentation of the journal in which they desire to be published, and (b) the meta-language of the genre

that they feel part of.[21] Consider, in this respect, the affinity between AJES and AJS on the one hand and between AER and JAMS on the other.

3. There is a specifically economic subject matter with distinctive properties, and authors inevitably need to use a specific language to discuss it. As we have seen, despite some differences between economic journals, economics is a linguistic genre of its own. This means that economics is *not* simply an amalgam of other scholarly discourse modes. For instance, published writings in AER cannot match the stringency of published writing in JAMS, while AJES lacks the empirical emphasis of sociological research in AJS.[22]
4. We have presented diachronic evidence that economics, as a scholarly discipline, may face a renaissance of the *methodenstreit*, although possibly one without quarrel, but only with divergence. Our linguistic data can be taken to suggest that the formal branch of economics—which, typically for economics, is ambiguous between abstraction and empirical-practical application—and the socio-scientific, theoretical-reflexive branch are drifting away from each other. This is certainly a somewhat daring hypothesis, yet the postulation of a growing disciplinary divide is not implausible given our findings.

It is fair to say that some of these findings are not entirely surprising. Rather, they confirm many expectations with regard to a subjectively felt increasing methodological gap. Still, we think that it is the merit of our data-driven, linguistic method of analysis to be able to back up this feeling with sound empirical facts. Thus, our findings can serve as a departing point for a more detailed analysis of developments within the community of economic authors and for a comparison to other sciences.

Finally, the existence of an intra-economic language community and the intradisciplinary alignment of economic journals depending on their appeal is reminiscent of Max Weber's description of language: "A common language, created by common tradition in one's family and one's social environment, eases mutual understanding and hence the emergence of social relationships [. . .]. But *per se*, a common

language does not imply communization. It only implies facilitation of exchange within such groups" (Weber 1922: 21–22).

Notes

1. Striking and often entertaining examples of this tradition are the early volume by Klamer, McCloskey, and Robert M. Solow (1988), Klamer's (1990) examination of Samuelson's different textbook editions, Bill Gerrad's (1991) interpretation of the interpretations of Keynes' *General Theory*, and the collected volume on *Economics and Language* edited by Willie Henderson, Tony Dudley-Evans, and Roger Backhouse (Henderson et al. 1993)—therein especially Tony Dudley-Evans' linguistic report of Friedman's theory—Bergeron and Pietrykowski (1999) on genre difference in economic literature (see this study for more references), as well as various contributions in Cullenberg et al. (2001). Other representatives of this tradition include Meriel and Thomas Bloor's essay on how economists modify knowledge claims, Robert S. Goldfarb's (1995) empirical study of plausible inferences in economics, Mark E. Blum's (1996) linguistic analysis of Sombart's work, and Waren S. Samuels's (2001) essay on the use of language in economics.

2. To our knowledge, Georg von der Gabelentz (1891) was the first to suggest that there are two powerful and competing tendencies in language usage, one toward ease of production (for speakers), and one toward ease of comprehension (for listeners).

3. While there exists a quite extensive linguistic literature on academic discourse in general, "there is very little on disciplines such as economics and sociology" (Swales 1990: 133). Worse yet, the few studies that have taken an empirical approach to economic scholarly writing from a linguistic point of view have not addressed those issues that are of interest to the present study. The studies that are parallel to ours and apply similar methods to other fields are Tinberg (1988) and Lindeberg (1994, 1996). Tinberg (1988) conducted a structural analysis of a handful of mainstream and neo-Marxist economic articles, finding that the neo–Marxist papers devoted substantially more space to model building than did the orthodox papers. Similarly, Lindeberg (1994, 1996) discussed introductory sections in economic, management, and business papers. Most similar to the present study is perhaps Chevalier and Hudson's (2001) study, in which they conduct a lexical analysis of intentional language in scientific articles in finance, albeit with a much smaller database than ours and a considerably more limited set of research questions. Mention should also be made of Conrad (1996), who uses a corpus-based approach comparable to ours. Her subject matter, though, is research articles and textbooks in biology. Also, although the present study will focus on an empirical analysis, it is warranted by the subject matter to pay attention to

studies dealing with the historical development of economics and attempting a narrative integration of economic, social, and political history. See, for example, the seminal work by Mirowski (1989, 2002) and Mirowski, ed. (1994) as well as historical studies on the rising use of mathematics in economics, such as Weintraub (2002), Weintraub, ed. (1992), Morgan (1990), Morgan and Morrison, eds. (1999), and Leonard (1994, 1995). For a history of economic thought perspective, see Hodgson (2001).

4. This is because in the period from 1990 to 2000, no substantially or statistically significant changes in economic academic prose were observable.

5. While different text types—for instance, research articles and book reviews—differ in their form, they do certainly not differ much in their orientation as regards content and/or ideology, given that they are published in the same journal.

6. For an overview, see Biber and Conrad (2003). For studies applying the method, see Biber (1986, 1988), Biber et al. (1998), Biber and Finegan (1989, 1994), and Atkinson (1992, 1996). For an approach to tracing diachronic developments similar to ours, see Biber and Finegan (1997).

7. Here and in the following, entries will have to be understood as linguistic lemmas; for instance, the lemma ASSUME subsumes all the formal lexical variations which may apply, such as *assuming, assumes*, or *assumed*.

8. See Biber (1988) for an outline of the use of factor analyses in discourse analysis. For an explanation of the term *frequency,* see below. For more detail on the factor analysis, see the Appendix.

9. We are interested here in the normative meaning of JUSTICE, viz. the principle or ideal of just dealing or right action. Therefore, frequencies were corrected manually for other uses of JUSTICE such as *criminal justice* or *Department of Justice*.

10. Note that the semantics of these modals is actually many times more complex than is presented here. This caveat notwithstanding, probability and authorial comment (also called epistemic semantic loading) appears to be their core meaning.

11. Note that as far as *we* is concerned, we are interested in the so-called exclusive *we* as opposed to the so-called inclusive *we*, which is why we manually corrected figures for inclusive *we*. In case of the former, authors refer exclusively to themselves, while in the case of the latter, they refer to themselves and their readership. (Example for inclusive *we*: "Combined with the j = 1 case of Lemma 3.1, we obtain Z.")

12. Probabilities of error (p) have been computed using a standard χ^2 test of independence (on the basis of nonstandardized frequencies) and refer to the likelihood that observed frequency differentials are due to sampling errors. Yates's correction for continuity has been employed where appropriate. Unless indicated otherwise, df = 1.

13. An exception may be the unusually high frequency of *divorce* in AER. The term, however, only occurs in Volumes 82 and 88 in our AER sample and is predominantly used in a rather technical sense (for instance, "In contrast, these data suggest that unilateral *divorce* has no significant impact on married women's labor-force participation unless the underlying marital-property laws in each state are considered," AER Vol. 88, 1998).

14. TTRs are defined as the ratio of the total number of different words (*types*) to the total number of words (*tokens*) in any given text. For instance, if a text contains in all 1,000 words, 300 of which are different, it will yield a TTR of 30percent. Because, however, TTRs vary wildly with text lengths—a text of 10 words is likely to contain close to 10 different words, while it is highly unlikely for a text of 10,000,000 words to contain 10,000,000 different words—we will standardize TTRs in this study by computing a TTR for every 1,000 words of running text and then calculating a running average.

15. This, of course, refers to linguistic complexity only. AER arguments are likely to be more complex mathematically.

16. According to the *Oxford English Dictionary*, the earliest clear use of the word dates from 1697, when it was used in a physics context ("The Fluids, pressing equally and easily yielding to each other, soon restore the Æquilibrium"). For the integration of the concept of equilibrium into economic thinking, see Quesnay (1767), especially p. 645. For a more recent discussion, see Ingrao and Israel (1990) and Mirowski (1989), especially pp. 238–241.

17. See http://khnt.hit.uib.no/icame/manuals/index.htm for more information on the LOB and F-LOB corpora.

18. In the linguistic reference corporas LOB and F-LOB, by contrast, social science terminology—and especially CULTURE/CULTURAL, HISTORY/HISTORICAL, and SOCIETY—increased in frequency by roughly 70percent from the 1960s to the 1990s. The sharp frequency decrease of this terminology in AER and the only slight increase in EJ and AJES, therefore, merit particular attention.

19. This is also observable in our reference corpora, where METHOD likewise decreased significantly in frequency over time.

20. Remember that we have seen earlier that JOM contains less empirical terminology than AJS, although it would certainly be hard to argue that microscopy is less empirical than sociology.

21. For a discussion of the "pressures for conformity," see, for example, Hodgson and Rothman (1999), Fölster (1995), and Stigler et al. (1995).

22. Readers with some background in linguistic theory will have noticed that these three points dovetail nicely with Karl Bühler's *organon* theory of language (Bühler, 1934), according to which the three main functions of language are *representation*, *expression*, and *appeal*.

References

Atkinson, Dwight. (1992). "The Evolution of Medical Research Writing from 1735 to 1985: The Case of the Edinburgh Medical Journal." *Applied Linguistics* 13: 337–374.

——. (1996). "The Philosophical Transactions of the Royal Society of London, 1675–1975: A Sociohistorical Discourse Analysis." *Language in Society* 25: 333–371.

Bergeron, Suzanne, and Bruce Pietrykowski. (1999). "Can There Be Genre Difference in Economic Literature?" In *What do Economists Know? New Economics of Knowledge*. Ed. Robert Garnett. London: Routledge.

Bühler, Karl. (1934). *Sprachtheorie: die Darstellungsfunktion der Sprache*. Jena: Fischer.

Biber, Douglas. (1986). "Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings." *Language* 62: 384–414.

——. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, Douglas, and Susan Conrad. (2003). "Register Variation: A Corpus Approach." In *The Handbook of Discourse Analysis*. Eds. Deborah Schiffrin, Deborah Tannen, and Heidi Hamilton. Malden, MA: Blackwell.

Biber, Douglas, Susan Conrad and Randi Reppen. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Biber, Douglas, and Edward Finegan. (1989). "Drift and the Evolution of English Style: A History of Three Genres." *Language* 65: 487–517.

——. (1994). "Multidimensional Analyses of Authors' Styles: Some Case Studies from the Eighteenth Century." In *Research in Humanities Computing*, Vol. 3. Eds. Don Ross, and Dan Brink. Oxford: Oxford University Press.

——. (1997). "Diachronic Relations among Speech-Based and Written Registers in English." In *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*. Eds. Terttu Nevalainen and Leena Kahlas-Tarkka. Helsinki: Societe Neophilologique.

Blum, Mark. (1996). "Breaks or Continuity in Sombart's Work: A Linguistic Analysis." In *Werner Sombart (1863–1941)—Social Scientist. Volume III: There and Now*. Ed. Jürgen Backhaus. Marburg: Meropolis.

Chevalier, Gisèle, and Richard Hudson. (2001). "The Use of Intentional Language in Scientific Articles in Finance." *Journal of Economic Methodology* 8: 203–228.

Conrad, Susan. (1996). "Investigating Academic Texts with Corpus-Based Techniques." *Linguistics and Education* 8: 299–326.

Cullenberg, Stephen, Jack Amariglio, and David Ruccio (eds). (2001). *Postmodernism, Economics and Knowledge*. London: Routledge.

Fölster, Stefan. (1995). "The Perils of Peer Review in Economics And Other Sciences." *Journal of Evolutionary Economics* 5(1): 43–57.

Gabelentz, Georg von der. (1891). *Die Sprachwissenschaft: Ihre Aufgaben, Methoden, und bisherigen Ergebnisse*. Leipzig: Weidel.

Gerrad, Bill. (1991). "Keynes's General Theory: Interpreting the Interpretations." *Economic Journal* 101: 276–287.

Goldfarb, Robert. (1995). "The Economist-as-Audience Needs a Methodology of Plausible Inference." *Journal of Economic Methodology* 2: 201–222.

Henderson, Willie, Tony Dudley-Evans, and Roger Backhouse. (1993). *Economics and Language*. London: Routledge.

Hodgson, Geoffrey. (2001). *How Economics Forgot History: The Problem of Historical Specificity in Social Science*. London: Routledge.

Hodgson, Geoffrey, and Harry Rothman. (1999). "The Editors and Authors of Economics Journals: A Case of Institutional Oligopoly?" *Economic Journal* 109: F165–F186.

Ingrao, Bruna, and Giorgio Israel. (1990). *The Invisible Hand: Economic Equilibrium in the History of Science*. London: MIT Press.

Klamer, Arjo. (1984). *Conversations with Economists: New Classical Economists and Opponents Speak Out on the Current Controversy in Macroeconomics*. Totowa, NJ: Rowman and Allanheld.

——. (1990). "The Textbook Presentation of Economic Discourse." In *Economics as Discourse: An Analysis of Language in Economics*. Ed. Warren Samuels. Boston: Kluwer.

——. (2001). "Making Sense of Economists: From Falsification to Rhetoric and Beyond." *Journal of Economic Methodology* 8: 69–75.

Klamer, Arjo, Deirdre McCloskey, and Robert Solow. (1988). *The Consequences of Economic Rhetoric*. Cambridge: Cambridge University Press.

Leonard, Robert. (1994). "Reading Nash, Reading Cournot: The Creations and Stabilisation of the Nash Equilibrium." *Economic Journal* 104: 492–511.

——. (1995). "From Parlor Games to Social Science: Von Neumann, Morgenstern, and the Creation of Game Theory, 1928–1994." *Journal of Economic Literature* 33: 730–761.

Levinson, Stephen. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature. Language, Speech, and Communication*. Cambridge: MIT Press.

Lindeberg, Ann-Charlotte. (1994). "Rhetorical Conventions in Scholarly Articles in Economics and Business Sciences: A Study of Introductions with a Special Reference to Knowledge Claims." In *Writing vs. Speaking: Language, Text, Discourse, Communication*. Eds. Svetla Cmejrkova, Frantisek Danes, and Eva Havlova. Tübingen: Narr.

——. (1996). "Argumentative Structures and Signals in Three Disciplines: Finance, Management, and Marketing." In *Theories, Models and Methodology in Writing Research*. Eds. Gert Rijlaarsdam, Huub van den Bergh, and Michel Couzijn. Amsterdam: Amsterdam University Press.

Mäki, Uskali. (1988). "How to Combine Rhetoric and Realism in the Methodology of Economics." *Economics and Philosophy* 5: 89–109.

——. (1995). "Diagnosing McCloskey." *Journal of Economic Literature* 33: 1300–1318.

——. (2000). "Performance against Dialogue, or Answering and Really Answering: A Participant Observer's Reflections on the McCloskey Conversation." *Journal of Economic Issues* 34: 43–59.

McCloskey, Deirdre. (1983). "The Rhetoric of Economics." *Journal of Economic Literature* 21: 481–517.

——. (1985a). "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests." *American Economic Review* 75: 201–205.

——. (1985b). *The Rhetoric of Economics*. Madison, WI: University of Wisconsin Press.

——. (1995). "Modern Epistomology against Analytic Philosophy: A Reply to Mäki." *Journal of Economic Literature* 33: 1319–1323.

——. (1998). "Two Replies and a Dialogue on the Rhetoric of Economics: Mäki, Rappaport, Rosenberg." *Economics and Philosophy* 4: 150–166.

——. (2001). "Stiglerite vs. Friedmanite Science." *Eastern Economic Journal* 27: 209–210.

McCloskey, Deirdre, and Stephen Ziliak. (1996). "The Standard Error of Regressions." *Journal of Economic Literature* 34: 97–114.

Mirowski, Philip. (1989). *More Heat than Light: Economics as Social Physics, Physics as Nature's Economics*. Cambridge: Cambridge University Press.

—— (ed). (1994). *Natural Images in Economic Thought*. Cambridge: Cambridge University Press.

——. (2002). *Machine Dreams: Economics Becomes a Cyborg Science*. Cambridge: Cambridge University Press.

Morgan, Mary. (1990). *The History of Econometric Ideas*. Cambridge: Cambridge University Press.

Morgan, Mary, and Margaret Morrison (eds). (1999). *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Moss, Laurence. (2002). "Editor's Introduction." *American Journal of Economics and Sociology* 61: 1–9.

Quesnay, François. (1767). "Ephémérides du Citoyen ou Bibliothèque Raisonnée des Sciences Morales et Politiques." In *Œvres Économiques et Philosophiques. Publiées avec une Introduction et des Notes par Auguste Oncken, 1888*. Ed. François Quesnay. Paris: Peelman.

Rayson, Paul, Geoffrey Leech, and Mary Hodges. (1997). "Social Differentiation in the Use of English Vocabulary: Some Analyses of the Conversational Component of the British National Corpus." *International Journal of Corpus Linguistics* 2: 133–52.

Rubinstein, Ariel. (2000). *Economics and Language: Five Essays*. New York: Cambridge University Press.

Samuels, Warren. (2001). "Some Problems in the Use of Language in Economics." *Review of Political Economy* 13: 91–100.

Stigler, George, Stephen Stigler, and Claire Friedland. (1995). "The Journals of Economics." *Journal of Political Economy* 105: 331–359.

Swales, John. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Tinberg, Jon. (1988). "The pH of a Volatile Genre." *English for Specific Purposes* 7: 205–212.

Weber, Max. (1922). *Wirtschaft und Gesellschaft. Grundriss der verstehenden Soziologie*. 5th edition. Tübingen: Mohr Siebeck.

Weintraub, Eliot (ed). (1992). *Toward a History of Game Theory*. Durham, NC: Duke University Press.

——. (2002). *How Economics Became A Mathematical Science*. Durham, NC: Duke University Press.

Appendix

Table 9

Explanatory Power of Factors (Extraction Method: Principal Component Analysis) 1

		Initial Eigenvalues	
Factor	Total	Percent of Variance	Cumulative Percent
1	11.571	60.898	60.898
2	5.034	26.496	87.394
3	1.556	8.188	95.582

Table 10

Factor Loadings (Extraction Method: Principal Component Analysis; Rotation Method: Varimax with Kaiser Normalization)

	Factor		
	1	2	3
THEORY/THEORETICAL	0.596	0.777	–0.162
MODEL	0.979	0.096	0.104
ASSUME/ASSUMPTION	0.945	0.146	0.273
OBSERVE/OBSERVATION	0.936	0.111	0.214
METHODOLOGY/METHODOLOGICAL	0.589	0.716	–0.266
METHOD/METHODICAL	0.661	0.421	–0.436
HISTORY/HISTORICAL	0.361	0.886	–0.239
CULTURE/CULTURAL	–0.045	0.976	–0.119
RATIONALITY/RATIONAL	0.473	0.877	–0.006
JUSTICE	–0.053	0.982	–0.051
SOCIETY	0.090	0.985	–0.126
THEOREM	0.107	–0.230	0.938
PROOF	0.289	–0.218	0.902
EVIDENCE	0.864	0.271	–0.218
DATA	0.943	0.324	–0.034
EQUILIBRIUM	0.967	–0.051	0.185
RESULT	0.930	0.289	0.152
EFFECT	0.969	0.234	–0.012
INSTITUTION/INSTITUTIONAL	0.422	0.832	–0.285