Soc Choice Welfare (1996) 13: 183–200

Social Choice and Welfare

The nature and measurement of freedom

Patrick Suppes

Stanford University, CSLI Ventura Hall Stanford, CA 94305–4115, USA

Received: 13 January 1995/Accepted: 6 November 1995

Abstract. This article develops an ergodic theory of freedom and its measurement. The key idea is that freedom and uncertainty are inextricably entwined. Measure-theoretic isomorphism of stationary processes is a natural way of identifying at an abstract level identity of the structure of uncertainty, which can be measured by the entropy rate of a stationary process. Various empirical examples of markets and elections are considered, as well as the Kantian, as opposed to utilitarian, philosophical framework of the general ideas.

1. Kantian prelude

Much traditional and modern moral philosophy does not give a central place to the concept of freedom, but this is not true of perhaps the currently most influential moral philosopher, Immanuel Kant. The concept of freedom occupies a central place in Kant's critical philosophy. Those who have read only his moral philosophy will think immediately of his concept of practical freedom. Those who have also studied the *Critique of Pure Reason* (1781/1787) will think of transcendental freedom. Freedom in this sense is the power of beginning an act spontaneously. It cannot be determined by another cause preceding it in time, according to the laws of nature. In the analysis in the Third Antinomy, formulated rather late in the *Critique*, Kant also examines the idea of theoretical freedom, i.e., the spontaneous power of beginning a process without casual predecessors in nature, but he settles for the complete determinism of the laws of nature. In the *Critique of Judgment* (1790) Kant develops in Part II a biological teleology whose concepts are justifiably used in the study of nature when the available resources of mechanical explanation are limited, i.e., more explicitly, our powers of giving such explanations are limited. For Kant such a teleology is scientifically useful for explaining natural products, but even more central is that it provides the link between our actions in the world of experience and the moral law that is the formal condition for employing our freedom.

I cannot develop these Kantian themes further here, but I sketch them to show that the underlying conception of the nature of freedom defended here is not derived from the utilitarian principles so commonly used by economists in their analysis of preference or choice. There are three fundamental concepts that are combined to describe the nature of freedom in the approach developed: individual autonomy, uncertainty and diversity of choice. I am under no illusion that these concepts, or the way I use them, will be adequate to account for all aspects of our intuitive ideas about freedom. I do think they provide a methodology for quantitatively analyzing some of the features of freedom most discussed in economics and political science, namely, the freedom of markets and elections. And this is what I mainly concentrate on in this article.

The central concept is the Kantian one of individual autonomy, i.e., of each rational being always thought of as an end in itself. Personal actions fully controlled by another, or by institutions, are not free actions. Intuitively, then, an election or market where there is no choice of candidates or products is not free. I propose, as the other two concepts, uncertainty and diversity, with entropy as a standard measure of uncertainty. The close connection between freedom and uncertainty is the main focus of this article. Entropy as the measurement of freedom is the subject of Sects. 2 and 3. The deeper reasons, derived from ergodic theory, for using this particular measure of uncertainty are developed there. The concept of diversity of candidates or products is the opposite of the widely studied psychological concept of similarity, and is the subject of Sect. 4, but connections to the psychological literature need a more thorough examination than is possible here. The final

Table 1. Entropy, in terms of popular and electoral votes, of American presidential elections. Data taken from Series Y, 79–83, *Electoral and Popular Vote Cast for President by Political Party: 1789–1968*, U.S. Bureau of the Census, 1975, pp. 1073–1074. Data for 1972–1992 taken from *Congressional Quarterly's Guide to U.S. Elections.*

Year	Popular	Electoral	Year	Popular	Electoral
1992	1.56	0.90	1888	1.22	0.98
1988	1.08	0.75	1884	1.21	0.99
1984	1.04	0.16	1880	1.19	0.98
1980	1 39	0.44	1876	1 07	1 00
1976	1.14	1.01	1872	1.03	1.12
1972	1.09	0.22	1868	1.00	1.16
1968	1.47	1.30	1864	0.99	1.15
1964	0.98	0.46	1860	1.87	1.50
1960	1.03	0.98	1856	1.52	1.12
1956	1.02	0.58	1852	1.23	0.59
1952	1.04	0.65	1848	1.37	0.99
1948	1.33	1.27	1844	1.14	0.96
1944	1.04	0.69	1840	1.00	0.73
1940	1.04	0.62	1836	1.00	1.65
1936	1.14	0.11	1832	0.99	1.10
1932	1.17	0.50	1828	0 99	0.90
1928	1.05	0.64	1824	1.82	1.88
1924	1.48	0.98	1820		0.14
1920	1.28	0.79	1816		0.75
1916	1.26	1.00	1812		1.01
1912	1.87	0.76	1808		1.08
1908	1.33	0.92	1804		0.40
1904	1.34	0.87	1800		2.02
1900	1.20	0.93	1796		2.73
1896	1.17	0.97	1792		1.60
1892	1.48	1.17	1789		2.47

section is a brief philosophical postscript concerned with re-examining individual autonomy and reformulating what Kant has to say about freedom, as well as expressing some skepticism about radical libertarianism.

2. Entropy as a measure of freedom

Entropy as a proposed measurement of freedom is phenomenological and result, rather than procedurally, oriented. Consider two elections. The first, E_1, has three candidates and each receives about $\frac{1}{3}$ of the votes. The second, E_2, has two candidates and the winner of the two receives about $\frac{3}{4}$ of the votes. Almost all of us would agree, I think, that the results as such are evidence of E_1 being more free than E_2. In saying this we are assuming the usual *ceteris paribus* conditions. Moreover, in matters political or economic there is a strong skeptical tradition that looks to results rather than intentions in judging the character of an institution or procedure.

I propose that we measure the freedom of a set A of alternatives by the entropy H of the actual chosen proportions, or relative frequencies, of the various alternatives, that is,

$$H(A) = -\sum_{i \in A} p_i \log p_i,$$

where log is to the base 2, $p_i \geq 0$ and if $p_i = 0$ then $0 \log 0 = 0$.

To give a feeling for the numbers, so to speak, Table 1 shows the entropy of American presidential elections. According to the measure proposed the elections with the maximum freedom, i.e., maximum entropy, of popular vote were those of 1860 and 1912. The tallies were as follows:

1860		1912	
Abraham Lincoln	1 865 593	Woodrow Wilson	6 296 547
J.C. Breckinridge	848 356	Theodore Roosevelt	4 118 571
Stephen A. Douglas	1 382 713	William H. Taft	3 486 720
John Bell	592 906	Eugene V. Debs	900 672
		Eugene W. Chafin	206 275
		Arthur E. Reimer	28 750

In contrast, the least free election as measured by popular vote was in 1964, with a measure of 0.98.

1964			
Lyndon B. Johnson	43 129 566	Clifton DeBerry	32 720
Barry M Goldwater	27 178 188	E Harold Muun	23 267
Eric Hass	45 219		

I will not examine the electoral college votes in detail, but note two things. The maximum entropy in either column is for the electoral vote in 1796, with a measure of 2.73. (The corresponding popular vote is not recorded, as is the case generally before 1824.)

Electoral vote, 1796			
John Adams	71	John Jay	5
Thomas Jefferson	68	James Iredell	3
Thomas Pinckney	59	George Washington	2
Aaron Burr	30	John Henry	2
Samuel Adams	15	S. Johnston	2
Oliver Ellsworth	11	C.C. Pinckney	1
George Clinton	7		

In 1824, the electoral vote has an entropy that is larger than that of any subsequent election, and in that year, because no candidate received a majority in the electoral college, the election was decided by the House of Representatives.

Electoral vote, 1824			
John Q Adams	84	Henry Clay	37
Andrew Jackson	99	W.H. Crawford	41

Table 2 analyzes the share of total value added by manufacture accounted for by the 200 largest U.S. manufacturing companies from 1947 to 1970. The analysis here is conditional. We take the percent of total value added by manufacture conditionalized on that contributed by the largest 200 companies, and then look at the contribution of the largest 50 companies, then the next largest 50, then the next largest after that, and finally the fourth group of 50. The entropy of this market, as conditionalized in terms of the 200 largest manufacturing companies, is an intuitive partial measure of the freedom of the market. If, for example, a single large company dominated the entire market the entropy would be 0. This table shows that there has been little change in freedom of the overall manufacturing market in the United States, at least among the 200 largest companies, from 1947 to 1970.

The second thing that the table shows is that the analysis of freedom in many cases will be conditional in form. But just as conditional probabilities are themselves probabilities so are conditional proportions that can be treated as relative frequencies or probabilities from the standpoint of entropy.

Here is another example of a different sort taken from Aydelotte (1963). In the vote on the corn laws in 1843 in the British House of Commons 127 liberals voted for, 52 against, 0 conservatives voted for and 331 against, and the measure was defeated. Three years later in 1846 when the measure passed, the vote was, on the

Table 2. Freedom in the market for manufacture in the United States based on data for 200 largest manufacturing companies, 1947–1970[1]

	1947	1954	1958	1962	1966	1970
Largest 50	17/39	23/37	23/38	24/40	25/42	24/43
Second 50	6/30	7/	7/	8/	8/	9/
Third 50	4/30	4/	5/	4/	5/	5/
Fourth 50	3/30	3/	3/	4/	4/	5/
Entropy	1.65	1.52	1.56	1.57	1.59	1.66

[1] Data taken from Series P 177–180, *Electoral and Popular Vote Cast for President by Political Party: 1789–1968*, U.S. Bureau of the Census, 1975, p 686

part of the liberals, 235 for, 10 against, and in the case of the conservatives, 114 for and 241 against. The entropy in the voting increased from 1843 to 1846. The specific numbers are 0.81 in 1843 and 0.98 in 1846.

Before considering some additional systematic general concepts, there are two remarks I want to make in some detail that are related to the examples given thus far. First, the measure of freedom I am proposing is, as I said at the beginning, mainly phenomenological. There is no suggestion that the measure itself says very much about the causal factors producing the measure at a given time, or a change in the measure from one period to another, whether in an election or in a market. There is, surely, an utter pluralism of causes of changes in entropy. Above all, increases in freedom occur not necessarily because of the intentional actions of individuals focusing on problems of freedom, but often because of what Aristotle termed incidental causation. This means that their intentions were focused on something else, but out of those intentions arose a mixture of results from the actions of many individuals that increased or decreased the freedom of a given institution, or political or social procedure.

On the other hand, I would not suggest that all such changes are unintentional. There are certainly historical cases of demagogues and dictators closing down freedom of elections so that they become simply a one-party farce. There are also well-documented cases of individuals intending to reduce the freedom of markets by intentionally arranging for an oligopoly having a small number of firms, or in contrast, stringent pursuit of anti-trust measures to increase the freedom of a given market. The point that I want to make is that the entropy measure itself gives no hint as to whether the causal conditions leading to the particular quantitative result, or change in quantitative measure, were intentional or not in character. It is rather like measuring the speed of a body in mechanics, which in itself, because of its purely kinematical nature, gives no clue to the forces that have produced the given state of velocity of the body, forces, which in the cases of uniform velocity could be quite remote in time. The second remark is that it is clear for a given set of alternatives A, with cardinality n, that the maximum entropy, in terms of proportions choosing the various alternatives, is produced by the uniform distribution of proportions or probability, of the n alternatives. A graph of the entropy of the uniform distribution of n, with $n = 2$ to $n = 8$ is shown in Fig. 1. Note that for n a power of 2, i.e., $2^p = n$, for some integer p, then the entropy of the uniform distribution of n alternatives is just p. Generally, the entropy of the uniform distribution on n alternatives is $\log_2 n$.

If the measure I am proposing seems at all appropriate, one consequence is that a typical aspect of the rhetoric of freedom is brought into question. In general, it is certainly not the case that we always will want to prefer more freedom to less in relation to any widely-used social process or institution. For example, we would get the greatest freedom of election in those cases in which the number of candidates was very large and about the same number of votes were obtained by each. But we might feel that such an outcome would lead to great difficulties of governance, in particular, great concern about the stability of any political arrangements reached by parties, all of which possessed such a small part of the total vote. I return to this case later. Another kind of example occurs in the case of the voting on the British corn laws. The fact that the positive vote in 1846 was very much closer than the negative vote in 1843, and therefore the entropy measure evaluates the later as more free than the earlier vote is not necessarily something we all view positively.

I shall have more to say about this matter from another viewpoint in the discussion of processes extended over time. But it is already a good point at which

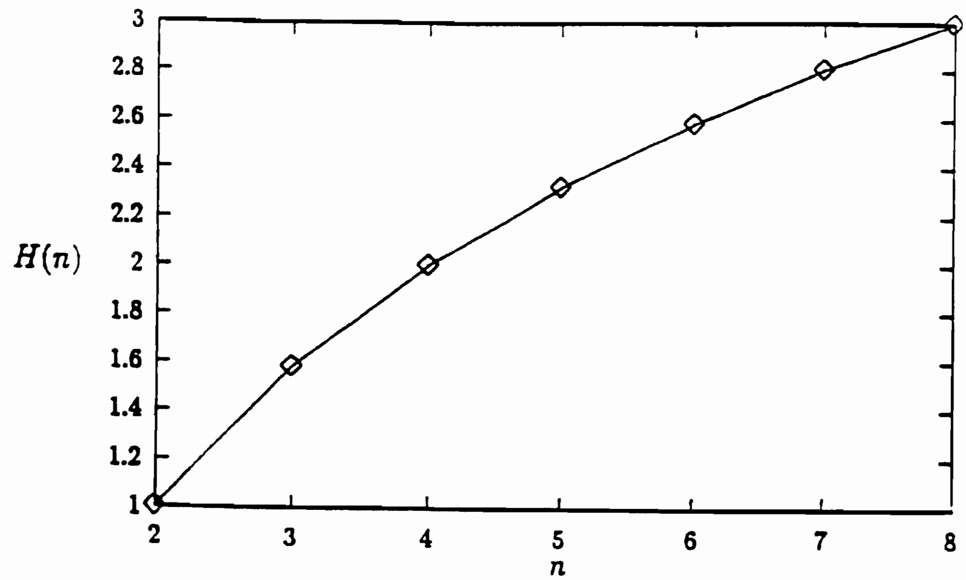

Fig. 1. Entropy for uniform distribution of n alternatives

to examine more carefully the viability of the proposed definition, given the somewhat counterintuitive results I have just mentioned. It may well be said by some political philosophers, but not by politicians or voters, that we do not really care about the outcomes of elections. What we care about are the political conditions under which they take place. If there is good evidence prior to the elections that there was a serious campaign among alternative candidates and individuals could freely state their political opinions, then what we judge as important are these conditions and not the fact that there was a real landslide of 90% to 10% in the actual voting. In this sense, it would be argued, the entropy measure is inappropriate. There is something in this criticism. It means that the analysis of freedom should be displaced from the results of the election to the procedures or processes leading up to it. We should then attempt to measure the presence of genuine dissent in the political dialogue preceding the election, the opportunities for choosing in terms of external social and political pressures, the resources available to the various candidates, etc. In my own view the outcome of this investigation would be in most cases fairly consistent with the analysis of the election results. Moreover, it is difficult to get quantitative and objective data about much of the political process leading to elections, but assuming the elections are themselves not dishonestly run, excellent quantitative data can be found in the results alone.

When there is freedom in the sense of entropy as measured quantitatively and as proposed here, it would be surprising to have a high measure of freedom for the process and a low one for the result. Notice, of course, that it is part of the rhetoric of politics that many people would say, even when very few resources were available, that it is still the case that individuals under the law were free to speak their minds about the candidates and to campaign as they wished in favor of whomever they wished. This is an important aspect of freedom and one that may not be satisfactorily caught by the measure I am proposing, but it is also one that is a source of skepticism about a political process that permits the kind of freedom just described and yet produces almost no results to back it up.

That the kind of increase in choices as reflected in actual exercise of that choice argues for an increase in freedom, is stated in vivid terms but in a wholly qualitative way in a letter of Verdi to the librettist of "Il Trovatore". He has this to say about how wonderful it would be to be able to break with the formal conventions of opera:

> If in operas there were no more cavatinas, no more duets, no more trios ... if the whole opera were one single piece, I would find that more reasonable and right... . It would be a good thing if, in the beginning of this opera, the chorus could be left out (every opera begins with a chorus); if Leonora's cavatina could be left out; and we begin right off with the Troubadour's song.

In a more mundane but systematic way, in a previous paper (Suppes 1987) in which I attempted to give axioms for the analysis of freedom of decision on the part of an individual, a critical axiom was that in a given situation an individual's freedom was always increased when the set of choices available to him was strictly increased. This axiom has been criticized as not meaning too much when the choices are simply formally available, but there is no evidence that they represent a genuine possibility of diversity for the chooser. The criticism has especially centered on the fact that the axioms gave no account of the concept of diversity of choices available to a given individual. I turn to this matter in Sect. 4. (For an excellent recent detailed study of the size of choice sets and ranking opportunity sets in terms of freedom, with extensive references to the literature, see Bossert et al. 1994, but there is no analysis of uncertainty and freedom.)

3. Stochastic freedom

A stochastic process χ is an indexed family $\{X_n\}$ of random variables. The index, discrete or continuous, is usually interpreted as time, and so it will be here. For simplicity and without any real conceptual loss, I consider only the discrete case with $n = 1, 2, 3, \ldots$, although some remarks will concern the doubly infinite case, $n = \ldots, -2, -1, 0, 1, 2, \ldots$. The usual assumption about the collection of joint probability distributions of any finite subsequence of the random variables being consistent is made.

The appropriate concept of entropy for a stochastic process χ is that of *entropy rate* $H(\chi)$ defined as follows

$$H(\chi) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n),$$

provided the limit exists. (Notice that $H(X_1, \ldots, X_n)$ is just the entropy of the first n random variables. We convert to a rate by dividing by n.)

A (discrete, finite) Bernoulli process is a stochastic process that is a sequence $X_1, X_2, \ldots,$ or possibly a doubly infinite sequence, with the X_n's independent and identically distributed random variables with a fixed finite range of values. It is easy to show that such a Bernoulli process χ has entropy rate

$$\begin{aligned} H(\chi) &= \lim_{n\to\infty} \frac{H(X_1, X_2, \ldots, X_n)}{n} = \frac{nH(X_1)}{n} H(X_1) \\ &= -\sum p_i \log p_i. \end{aligned}$$

We take the measure of freedom to be the entropy rate of the process.

Consider a market over time in which m individuals are sellers and n are buyers and at each period each buyer makes a purchase from exactly one seller. As before, the uniform probability distribution on the set of m^n possible transactions would define a discrete (and finite-valued) Bernoulli process, which would be for m^n possible transactions the stochastic process with maximum entropy rate and thus the one of this size with maximum freedom.

I simplify the analysis at this point by considering only the sellers as the states of the market process. The probabilities of each of the m states, i.e., sellers, represents the probability a random buyer will choose that seller at the given time. In application of these ideas to market data we would often need to estimate $p_{i,n}$ for seller i at the end of time period n by the relative proportion of the market seller i had for that period and make no attempt to identify the behavior of individual buyers. This asymmetry in the treatment of buyers and sellers is common in the analysis of markets and correspondingly, in the case of elections for candidates and voters. However, it is to be emphasized that this limited kind of data analysis is not at all satisfactory for a study of market processes over time, when the entropy rate depends on the transition data for individual buyers, as will become clear in the sequel. I note here that a sample path for a buyer is the sequence of states occupied by the buyer from one time period to another, with the state representing the seller with whom the buyer has a transaction. Although I do not do it here, for actual data analysis it will be desirable to introduce a state corresponding to a buyer not making a transaction in a given time period. There is little doubt that most sellers would shudder at the utter randomness of a Bernoulli market from one period to the next, as would most candidates at a sequence of elections with a corresponding Bernoulli character. Many firms would accept, even if not maximally satisfied, a market that is about evenly divided among a relative small number of sellers, but would be aghast at the utter lack of customer loyalty as the buyers randomly shifted at each period from one seller to another.

The necessity of considering the time course of a market, and not just cross-section data, in evaluating its freedom can be well illustrated by a market with just three sellers. We can look at the three-state Markov market with the transition matrix

	1	2	3
1	$1-2\varepsilon$	ε	ε
2	ε	$1-2\varepsilon$	ε
3	ε	ε	$1-2\varepsilon$

The "freedom curve" or entropy rate as a function of ε for this Markov market is shown in Fig. 2. The ε that makes this market have the same entropy rate as the Bernoulli process $B(\frac{1}{2}, \frac{1}{2})$ with two alternatives of equal probability is between 0.10 and 0.11. (An explicit formula for the entropy rate of a Markov market is given below.)

In the spirit of representing games in extensive form, we can easily conceptualize a simple psychological process model to represent the three-state market. The graph is shown in Fig. 3 for a buyer who bought from seller 0 on trial n. On the first move, throw a biased die to decide whether to change seller or stay put. If change is the outcome, randomly choose one of the other two sellers.

I hasten to add that Fig. 3 does not really show why this three-seller market with ε about 0.10 has the same entropy as $B(\frac{1}{2}, \frac{1}{2})$. More generally, for a stationary

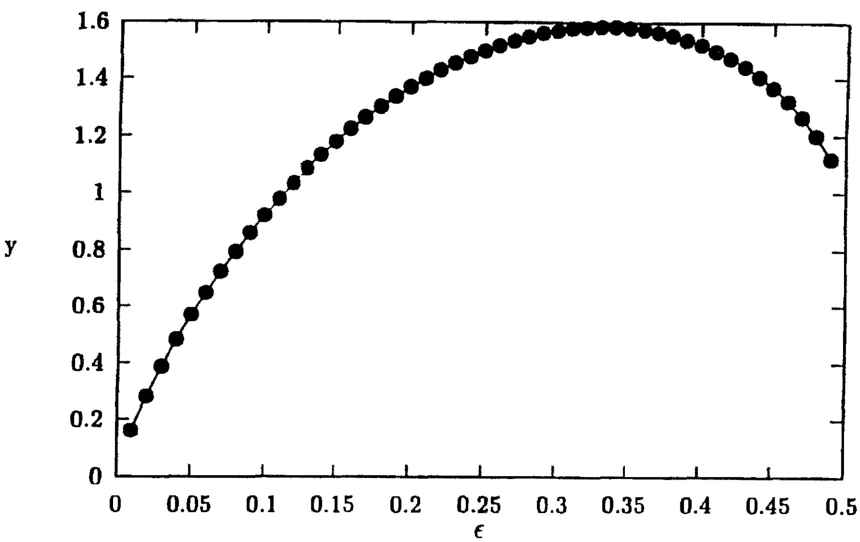

Fig. 2. Entropy rate for Markov market with three sellers as function of ε

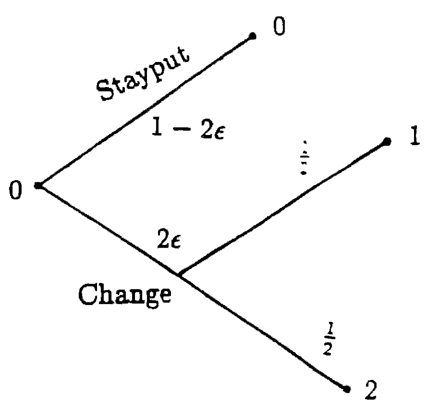

Fig. 3. Extensive form of market process

process the entropy rate as defined above is equal to the conditional entropy rate defined as

$$H'(\chi) = \lim_{n\to\infty} H(X_n|X_{n-1}, \ldots, X_1),$$

provided the limit exists, which it does for stationary processes. (For proof of the equality of the two definitions and an excellent discussion of the entropy of processes, see Cover and Thomas 1991, Ch. 4.) For a (first-order) stationary Markov process, as in our example,

$$\begin{aligned} H'(\chi) &= \lim H(X_n|X_{n-1}, \ldots, X_1) \\ &= H(X_2|X_1) \\ &= -\sum_x p(x) \sum_y p(y|x) \log p(y|x) \end{aligned}$$

and so it is easy to show for our Markov market example as defined above that as $\varepsilon \to 0$, $H(\chi) \to 0$. (Hereafter, I drop the distinction between H and H' in view of their equality for stationary processes.)

I now turn to the concept that is critical for making entropy rate the essential measure of the freedom of a market or election process – I add the word "process" to emphasize we are considering processes, not one-time cross-sections. The central question is this. How do two markets, or a market and an election, for that matter, compare in their intuitive sense of freedom if they have the same entropy, and contrariwise? As far as I know, this is not a question that has been previously addressed in economics or political science. There have been several prior uses of entropy to measure the one-time cross-section distribution of a market, as part of a more general consideration of indices of concentration (Encaoua and Jacquemin 1980, Curry and George 1983, Tirole 1988, Ch. 5, Foley 1994), but not of a market as a stochastic process. More importantly, entropy, as an invariant feature of what structural properties of stationary stochastic markets, has not been examined. The answer lies ready at hand in the mathematical literature on ergodic theory. In many cases of conceptual interest two stationary stochastic markets or elections will have the same entropy rate if and only if they are isomorphic in the measure-theoretic sense. It is this latter concept that needs to be defined.

Let us first begin with a standard probability space $(\Omega, \mathfrak{I}, P)$, where it is understood that $\mathfrak{I}$ is a σ-algebra of subsets of Ω and P is a σ-additive probability measure on $\mathfrak{I}$. We now consider a mapping T from Ω to Ω. We say that T is *measurable* if and only if whenever $A \in \mathfrak{I}$ then $T^{-1}A = \{\omega \colon T\omega \in A\} \in \mathfrak{I}$, and even more important, T is *measure preserving* if and only if $P(T^{-1}A) = P(A)$. T is *invertible* if the following three conditions hold: (i) T is 1–1, (ii) $T\Omega = \Omega$, and (iii) if $A \in \mathfrak{I}$ then $TA = \{T\omega \colon \omega \in A\} \in \mathfrak{I}$. In the application we are interested in, each ω in Ω is a doubly infinite sequence and T is the *right-shift* such that if for all n, $\omega_n = \omega'_{n+1}$ then $T(\omega) = \omega'$. Intuitively this property corresponds to stationarity of the process – a time shift does not affect the probability laws of the process, and we can then use T to describe orbits or sample paths in Ω.

We now characterize isomorphism of two probability spaces on each of which there is given a measure-preserving transformation, whose domain and range need only be subsets of measure one, to avoid uninteresting complications with sets of measure zero that are subsets of Ω or Ω'. Thus we say $(\Omega, \mathfrak{I}, P, T)$ is *isomorphic in the measure-theoretic sense* to $(\Omega', \mathfrak{I}', P', T')$ if and only if there exists a function φ: $\Omega_0 \to \Omega'_0$, where $\Omega_0 \in \mathfrak{I}$, $\Omega'_0 \in \mathfrak{I}'$, $P(\Omega_0) = P(\Omega'_0) = 1$, and φ satisfies the following conditions:

(*i*) φ is 1–1,
(*ii*) If $A \subset \Omega_0$ and $A' = \varphi A$ then $A \in \mathfrak{I}$ iff $A' \in \mathfrak{I}'$,
and if $A \in \mathfrak{I}$

$$P(A) = P'(A'),$$

(*iii*) $T\Omega_0 \subseteq \Omega_0$ and $T'\Omega'_0 \subseteq \Omega'_0$,
(*iv*) For any ω in Ω_0

$$\varphi(T\omega) = T'\varphi(\omega).$$

I emphasize that the isomorphism in the measure-theoretic sense of two markets, two elections, or a market and an election seems at the right level of abstraction. The isomorphism expresses that the two structures have the same degree of uncertainty and thus the same structural freedom, even though they differ considerably in other characteristics. The fundamental point is that our conception of freedom needs to be at a rather high level of abstraction in order to be conceptually useful. It would be of little use if we ended up by making the freedom of each market or election *sui generis*, and thus not comparable to any other. What we should have is a methodology for comparing degrees of freedom. The isomorphism in a measure-theoretic sense of two stationary stochastic processes provides the important step of giving us a meaningful basis in terms of uncertainty for judging equivalence in freedom. Note why this is so. The φ function mapping one process into another is measure-preserving, so there is a structural isomorphism between corresponding events of the two processes such that they have the same probability. It is precisely the fact that the mapping carries events into events of the same probability that supports the claim that isomorphism in the measure-theoretic sense represents equivalence of freedom of markets or elections.

On the other hand, it is equally important to note that isomorphism in the measure-theoretic sense of two stochastic markets only means isomorphism in the structure of uncertainty, as I have called it. Such isomorphism does not imply observational equivalence, nor would we want it to. For example, a Bernoulli market and a Markov market with strong dependence from one period to the next can be isomorphic in the measure-theoretic sense but easily distinguishable by a chi-square test for dependence. What we want to be able to say about these two markets is that they are equivalent in terms of freedom, but clearly different in other respects.

To show how recent fundamental results are about the relation between entropy rate and measure-theoretic isomorphism, I note that it was an open question in the 1950s whether the two finite-state discrete Bernoulli processes $B(\frac{1}{2}, \frac{1}{2})$ and $B(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ are isomorphic. (The notation here should be clear, as explained earlier; $B(\frac{1}{2}, \frac{1}{2})$ means that the probability for the Bernoulli process with two outcomes on each trial is that for each trial the probability of one alternative is $\frac{1}{2}$ and of the other $\frac{1}{2}$.) The following theorem clarified the situation.

Theorem 1 (Kolmogorov 1958, 1959; Sinai 1959). *If two finite-state, discrete Bernoulli or Markov processes have different entropies, then they are not isomorphic in the measure-theoretic sense.*

Then the question became whether or not entropy is a complete invariant for measure-theoretic isomorphism. The following theorem was proved a few years later by Ornstein.

Theorem 2 (Ornstein, 1970). *If two finite-state, discrete Bernoulli processes have the same entropy rate then they are isomorphic in the measure-theoretic sense.*

This result was then soon easily extended.

Theorem 3 (Adler et al. 1972). *Any two irreducible, stationary finite-state discrete Markov processes are isomorphic in the measure-theoretic sense if and only if they have the same periodicity and the same entropy.*

We then obtain

Corollary 2. *An irreducible, stationary, finite-state discrete Markov process is isomorphic in the measure-theoretic sense to a finite-state discrete Bernoulli process of the same entropy rate if and only if the Markov process is aperiodic.*

Given a stationary stochastic market or election the case is a good one for accepting entropy rate as an appropriate measure of freedom. There are, however, two serious reservations. The first reservation concerns the problem of diversity, to which the next section is devoted. The second is the simple fact that there are in the real world no doubly infinite stationary stochastic markets or elections. To

Table 3. Transition matrix for soft-drink choices

$n/n+1$	Coke	7-up	Tab	Like	Pepsi	Sprite	D-Pep	Fresca	p_∞	Entr.
Coke	0.61	0.11	0.01	0.03	0.13	0.06	0.01	0.04	0.29	1.45
7-up	0.19	0.45	0.00	0.06	0.14	0.10	0.01	0.05	0.18	1.82
Tab	0.08	0.12	0.16	0.36	0.08	0.04	0.08	0.08	0.03	2.38
Like	0.09	0.15	0.09	0.15	0.24	0.04	0.13	0.11	0.06	2.55
Pepsi	0.18	0.13	0.01	0.03	0.51	0.07	0.03	0.04	0.23	1.68
Sprite	0.11	0.18	0.03	0.07	0.16	0.33	0.03	0.09	0.10	2.27
D-Pep	0.09	0.05	0.18	0.09	0.12	0.09	0.26	0.12	0.04	2.52
Fresca	0.22	0.09	0.05	0.11	0.15	0.11	0.07	0.20	0.07	2.37
							Total Entropy			1.85

take advantage of the intuitions and results of ergodic theory this rather drastic abstraction has been used, a practice not uncommon in economics, but not to be commended. It is a task for the future to modify the theoretical framework to make it more empirically realistic, but still able to deal with markets or elections as dynamic processes over an extended period of time, not just in terms of a single cross section. (What is critical is approximate stationarity, and fortunately this can be statistically evaluated for the finite sequence of time periods available.)

To illustrate more concretely how empirical analysis of entropy rate for market or election processes can be undertaken, even if the stationarity assumptions are only roughly approximated, I consider two market examples.

The first example uses transition data observed six times for 264 buyers choosing one of eight soft-drink brands (Bass 1974). The buyers serving as subjects in the experiment were required to select a 12-ounce can of soft drink four days a week for three weeks from among the eight brands shown in Table 3. All brands were available six of the twelve days. The estimated probability transition matrix shown in Table 3 represents the average of the five transition samples from the six days of complete observations. The stationary probabilities, for choice of each brand, which are the choice probabilities that would hold asymptotically for the given transition data, are shown to the right in the column labeled p_∞. The next column to the right shows the entropy of each row, i.e., $-\sum_j p_{ij} \log p_{ij}$. The total entropy rate for the market, which is just the p_∞-weighted average of the row entropies, is 1.85, as shown at the bottom right of the table.

In Table 4 I show the transition data on U.S. car purchases from the 1985 New Car Buyer Competitive Dynamics Survey of J.D. Powers and Associates. The data given here are from a data reduction used in McCarthy et al. (1992). The original data set consists of 30 142 automobile purchases in 1985, together with the record of the previous purchases of each buyer. McCarthy et al. (1992) drew a 25% random sample of 7523 observations, a small number of which were eliminated because of incomplete records. The first part shows the estimated probability transition matrix for the sample drawn. The second part shows the California subsample and the third the non-California subsample. As can be seen the purchases are divided into four categories: Chrysler, Ford, General Motors and Foreign, with obviously all purchases of foreign cars lumped together in the last category. The total entropy of the national sample is 1.84, that of the California sub market is 1.93, and that of the non-California sub market is 1.79.

It may seem surprising that the eight-brand soft drink market, as sampled, has almost exactly the same entropy rate as the four-brand analysis of the U.S. national

Table 4. Transition matrix for automobile purchases

National sample $s = 7523$						
	Chrysler	Ford	Motor	Foreign	p_∞	Entropy
Chrysler	0.27	0.19	0.38	0.16	0.27	1.92
Ford	0.28	0.30	0.32	0.10	0.18	1.89
Motor	0.27	0.14	0.49	0.10	0.39	1.74
Foreign	0.27	0.10	0.26	0.37	0.16	1.88
				Total Entropy		1.84
California subsample						
Chrysler	0.31	0.22	0.27	0.20	0.29	1.98
Ford	0.20	0.33	0.27	0.20	0.19	1.97
Motor	0.31	0.12	0.37	0.20	0.28	1.89
Foreign	0.31	0.12	0.19	0.38	0.24	1.88
				Total Entropy		1.93
Non-California Subsample						
Chrysler	0.27	0.18	0.39	0.16	0.25	1.91
Ford	0.27	0.36	0.29	0.08	0.18	1.85
Motor	0.23	0.13	0.55	0.09	0.43	1.66
Foreign	0.27	0.10	0.29	0.34	0.14	1.89
				Total Entropy		1.79

automobile market, but the explanation seems obvious. Coke and Pepsi dominate the soft-drink market as sampled in a way that is not the case for the automobile market, and here when I speak of *dominate* I refer to the whole process of repeated choice and the consequent transition matrix. The diagonal entropy of 0.61 for Coke is not matched in magnitude by any other estimated probability in the four matrices of Table 4. The closeness of the measure of entropy rate, and as proposed here, the measure of freedom of these markets has a sound theoretical basis in the measure-theoretic concept of isomorphism. The corresponding qualitative intuitive argument given above supports the same conclusion. A possible criticism in terms of the diversity of buyers is dealt with at the end of the next section.

4. Measurement of diversity

There are two puzzles that can motivate the treatment of diversity I consider here. The first, formulated by Ken Binmore at the 1994 Caen Conference where the first draft of this article was presented, concerns an election with two candidates that seem essentially identical from a political standpoint. The answer I propose here is relatively simple and needs further elaboration. But I think it catches the central idea of a reasonable solution. For a given person or group, let there be just m properties or characteristics that are relevant for a candidate in an election or product in a market. To keep things simple, let these properties be only qualitative ones – either they are or are not possessed by a candidate or product. No quantitative measure is admitted in the present version. Then for m relevant properties there are just 2^m possible types of candidates or products for the given situation. My second simplification is to ignore the reasonable claim that different relevant characteristics have different importance, and therefore weights should be introduced to measure relative importance. (I return to this matter of weights later.) Given these two limiting assumptions of having only qualitative properties and only equal weighting of importance, it is straightforward to change the definition of entropy to be for the proportions of types, not individuals, in a given election or market. I give the revised definition only for a fixed point in time, not for the entropy rate of a process, but it is clear the new definition extends to processes immediately. Let T be the set of types and $t = 2^m$, the number of types generated by m properties. Then the entropy $H(T)$ is defined in the expected way by summing over types:

$$H(T) = -\sum_{i=1}^{t} p_i \log p_i.$$

The form of the definition looks, of course, exactly the same. But now it is for types not individuals. We still have a single measure of freedom, but with a revision of the way the idea of entropy as a measure of uncertainty is applied.

Notice how directly this characterization of relevant types solves the first puzzle. The entropy of an election consisting of candidates who are the same with respect to the set of relevant characteristics or properties is zero, for there is just one relevant type, and the entropy of the election is zero.

The second puzzle is this. Given just the measurement of entropy as the measurement of freedom, why not consider the best election in the sense of freedom the one where each voter votes for himself, the result that obviously has the maximum cross-sectional entropy for a given population of voters. When types are introduced, the election with maximum entropy is just the one having a uniform distribution of votes for the relevant types, independent of the total number of voters. It is worth noting that there is another and different theoretical solution to this puzzle. Consider the stochastic process in which each voter always votes for himself. Then the transition matrix has 1's on the diagonal, 0's everywhere else, and the entropy rate is zero, so freedom is not maximized at all.

A weighted model of types is more realistic, because of the nearly uniform agreement that relevant political or economic properties, characteristics or issues vary in importance. Candidates' views on foreign policy matter more for almost everybody than views on the budget for national parks. As much as differences in color matter in choice of a car, large differences in price matter more. Let w_i be the weight assigned to relevant type i by an individual or group, normalized so that $\sum w_i = 1$ with $w_i \geq 0$ for all i. We can then define an entropy-like quantity

$$\begin{aligned}\mathscr{W}(T) &= -\sum_{i=1}^{t} p_i \log p_i(tw_i) \\ &= H(T) - E(\log W) - \log t\end{aligned}$$

which has the property that for the uniform distribution of weights, i.e., $w_i = 1/t$, we have $\mathscr{W}(T) = H(T)$, since for the uniform distribution

$$\mathscr{W}(T) = -\sum_{i=1}^{t} p_i \log p_i \left(t \cdot \frac{1}{t}\right) = H(T).$$

Note that $E(\log W)$ is just the expectation with respect to the probability distribution p_i of $\log W$, where W is the random variable with values w_i.

The move from products or goods to properties of products is a familiar one in the economic theory of consumer demand. It has been an explicit focus of theory at least since Lancaster (1966). The current theoretical developments are well presented in Beath and Katsoulacos (1991). But, as far as I know, in the theoretical literature on product differentiation there has been no concern to analyze the freedom of markets, especially markets as stochastic processes. On the other hand, connections are made explicitly in this literature between product differentiation and market imperfection, relations between product differentiation and firms' beliefs about the extent to which consumers are informed about the products in a given market, etc. Lack of application explicitly to analysis of freedom is also characteristic of the large psychological literature on similarity. Clearly the analysis given here can be much deepened by connecting in some detail the ideas developed to these two substantial lines of prior research.

4.1. Diversity of buyers and voters

Thus far the analysis of diversity has centered on the "objects" of choice, whether candidates in an election or market products. It is obvious, however, that corresponding questions can be raised about buyers and voters. The implications are different, but just as important as those for the objects of choice.

Underlying the ergodic concepts used here, as applied to stationary processes, is the condition that in principle the sample path of choices of each buyer or voter should replicate in time the features of the full process. This allows for no significant relevant differences between buyers, and is consequently too restrictive. An approach to the study of consumer behavior that is widely used is to divide consumers into classes forming submarkets. The most widely used division is that between "loyalists" and "shoppers", i.e., between those who have brand loyalty and those who shop for the best bargain. (Corresponding distinctions have been made for voters.)

The ergodic analysis developed here can then be applied to each of these submarkets separately. Obviously, for a given set of products shoppers operate in a submarket with greater freedom than do loyalists. Identification of finer divisions of many markets is an important pursuit for many firms, and consideration of the empirical basis for such further divisions is important for further development of the ergodic analysis begun here.

5. Final philosophical remarks

I mentioned in the Introduction I would return to Kant. The latter part of the *Critique of Pure Reason* is devoted to four antinomies that pure reason can generate. All four of the antinomies are about problems that have a long history in philosophy. The first concerns whether the world is finite or infinite in space and whether it has a beginning in time or the past is of infinite extent. The second concerns whether matter is continuous (Aristotle) or discrete (Democritus). The third concerns causality, and is my focus here. Do all phenomena have determinate causes, or are there some that do not? Although Kant concludes by adopting the deterministic view, in his proof of the contrary he introduces just the right theoretical concept of freedom, namely, as the property of a cause that is itself absolutely spontaneous. To eradicate the bugaboo of universal determinism that has plagued the discussions of freedom since ancient times, we need to accept Kant's Third Antinomy as genuine by recognizing there is no scientific basis to support universal determinism.

Additional modern support comes from results in ergodic theory. Isomorphism in the measure-theoretic sense is too weak, as it should be, for the reasons given earlier, to serve by itself as a basis for claiming the empirical indistinguishability of deterministic and indeterministic theories of many phenomena. But if we add to measure-theoretic isomorphism a concept of geometric congruence with a fixed finite accuracy, then for a variety of ergodic systems such indistinguishability can be proved. (For both historical and conceptual details, see Ornstein and Weiss 1991). Because billiards have served as a philosophical paradigm of a mechanical system, a vivid example is the motion of a single billiard ball on a table with a convex obstacle in the middle of the table, a case of what is called Sinai billiards, after the Russian mathematicism Ya. G. Sinai, who has studied such systems extensively. Making the usual idealizations, such as that the ball is a point particle and that perfect elasticity and reflection hold, we can show ergodic behavior on the part of the ball, construct an ergodic measure for the deterministic Newtonian mechanical theory of the ball's motion, and prove isomorphism in the sense of congruence with finite accuracy to a Markov stochastic model of the motion. For more detailed discussion, see Suppes (1993), where I argue for the transcendental character of determinism, and Suppes (1994) where I then go on to argue in a modern neo-Kantian way for the empirical character of freedom.

I mentioned at the beginning individual autonomy as a central aspect of freedom. I now want to argue for the close relation of such autonomy to the ergodic behavior of free markets and elections. The behavior of markets and elections is made up of the choices of many individuals. Moreover, it is a fundamental result of ergodic theory that the individual sample path of a stationary process reproduces the main features of the process. As the saying goes, for such processes the time averages equal the ensemble averages. This means that typical individuals exhibit the same degree of unpredictability as the process as a whole. We scarcely expect markets or elections to be Bernoulli processes, without any memory of the past, and the same is true of individuals. But, we also do not expect the entropy to be zero, and this also is true of individuals. The surest sign of individual autonomy in economic and political choices is their uncertainty, and thus unpredictability to some degree, even when focused only on the individual decision maker. Perhaps even more important, this "factor" of unpredictability in autonomous individual choice is present for either deterministic or stochastic models of behavior. Put another way, there are far reaching theoretical reasons for holding that inherently unpredictable features of autonomous individual choice cannot be eliminated by any deterministic psychological model of choice, no matter how appealing its concepts may seem. If we cannot successively predict the behavior of billiard balls, we are unlikely to do so for human behavior.

I want to end by setting the analysis of freedom I have embarked on in this article in a more general philosophical framework. My natural temperament is close to those libertarians who make freedom a principle *primus inter pares*. The kinds of results derived here reinforces skepticism toward such an attitude. They support, on the contrary, strong philosophical leanings toward pluralism in almost all matters of any importance. Libertarians who would make freedom above all the most important principle of social organization would be led – if my characterization of its measurement is at all correct – to advocate results that would often be too unstable both politically and socially, given the aim was to maximize the entropy of economic and social processes without other constraints. Quantitative analysis of freedom is seldom if ever given by single-minded libertarians. Even for them purely Bernoulli markets and elections would be too much.

The pluralistic approach to equity advocated in Suppes (1988) can be strengthened along the following lines. Economists, and, on some occasion, political theorists have been misled by mistaken conceptions of the way mathematical methods are used in physics. It is sometimes taken as gospel that physics operates from a small number of fixed principles and has as its claim to fame the derivation from these principles of predictions about the behavior of many different kinds of physical phenomena. The truth is more wayward and devious. Physics, like other parts of science, operates with a hodgepodge of principles. If you took 10 000 pages of some given section of *Physical Review*, let us say quantum optics, which is only a part of last year's publications in this area, you would find it impossible to reduce to a small set the principles used. The arguments given to analyze data from a theoretical standpoint or to propose new theoretical results that need experimental testing are dense and numerous. The glory of physics is the chaos of its organization of principles. The rhetoric of physicists often belies this fact, but critical examination of what physicists actually do both in their scientific talk and in their scientific writings easily supports what I have to say. A favorite example of mine is the notorious quote of Dirac to the effect that now that quantum mechanics has been developed, chemistry is just an applied branch of physics. It turns out that nothing could be further from the truth. The physical behavior of almost no interesting chemical structures can be computed from scratch from fundamental principles of quantum mechanics. Just as in the past, but even more so, chemists need insight and intuitive feeling for the particular areas in which they work to derive and use the particular results needed.

Why should we ever expect in economics, philosophy or political theory something overridingly simple about the organization of society in terms of choice, welfare, or freedom, given the complexity of the behavior of even the simplest objects. Just the same pluralism to be found in the hard natural sciences should be present in economics. In particular, utilitarianism, which does not dominate philosophical thinking about moral philosophy, should not play the dominant role it does in the thinking of economists about choice and welfare. Much utilitarian thinking is to be commended. But not if it is moved from a pluralistic context to a role of intellectual imperialism. The same I say with no more reluctance of libertarianism and too great an accent on freedom. Freedom is important but not overwhelmingly so, when quantitatively considered. As analyzed here, maximizing freedom entails maximizing uncertainty, and this would be too simple a goal for a person or society if it dominated all other considerations. On the other hand, the contrapositive inference that no uncertainty entails no freedom is to be taken even more seriously. The inevitable conclusion is that the tension between freedom and other personal or societal goals cannot be eliminated, and it is irrational to believe it is possible.

Acknowledgment. I thank Ken Binmore and Peter Hammond for their useful criticisms, which have improved both the analysis and exposition of the key ideas.

References

Adler RL, Shields P, Smorodinsky M (1972) Irreducible Markov shifts. Ann Math Statistics 43(3): 1027–1029

Aydelotte WO (1963) Voting patterns in the British House of Commons in the 1840s. Comp Stud Soc Hist V, 134–163

Bass FM (1974) The theory of stochastic preference and brand switching. J Marketing Res, XI: 1–20

Beath J, Katsoulacos Y (1991) The economic theory of product differentiation. Cambridge University Press, Cambridge

Bossert W, Pattanaik P, Xu Y (1994) Ranking opportunity sets: an axiomatic approach. J Econ Theory 63: 326–345

Congressional quarterly's guide to U.S. elections (1994) Congressional Quarterly Inc, Washington, D.C.

Cover TM, Thomas JA (1991) Elements of information theory. Wiley, New York

Curry B, George K (1983) Industrial concentration: a survey. J Ind Econ XXX: 203–255

Encaoua D, Jacquemin A (1980) Degree of monopoly, indices of concentration and threat of entry. Int Econ Rev 21: 87–105

Foley DK (1994) A statistical equilibrium theory of markets. J Econ Theory 62: 321–345

Friedman NA, Ornstein DS (1970) On isomorphism of weak Bernoulli transformations. Adv Math 5: 365–394

Kant I (1790) The critique of judgement (J. Creed Meredith, Trans. (1928/1952)). Oxford University Press, New York

Kant I (1781/1787) Critique of pure reason (N. Kemp Smith, Trans. (1929/1965)). St. Martin's Press, New York

Kolmogorov AN (1958) A new metric invariant of transient dynamical systems and automorphisms in Lebesgue space. Dokl Akad Nauk SSSR, 119: 861–864 (Russian) MR 21 #2035a

Kolmogorov AN (1959) Entropy per unit time as a metric invariant of automorphism. Dokl Akad Nauk SSSR, 124: 754–755 (Russian) MR 2 # 2035b

Lancaster K (1966) A new approach to consumer theory. J Polit Econ 74: 132–157

Lancaster K (1979) Variety, equity, and efficiency. Columbia University Press, New York

McCarthy PS, Kannan PK, Chandrasekharan R, Wright GP (1992) Estimating loyalty and switching with an application to the automobile market. Manage Science, 38(10): 1371–1393

Ornstein DS (1970) Bernoulli shifts with the same entropy are isomorphic. Adv Math 4: 337–352

Ornstein DS, Weiss B (1991) Statistical properties of chaotic systems. Bull (New Series) Am Math Soc 24(1): 11–116

Sinai YaG (1959) On the notion of entropy of a dynamical system. Dokl Akad Nauk SSSR 124: 768–771

Suppes P (1987) Maximizing freedom of decision: An axiomatic analysis. In: Feiwel GR (ed.) Arrow and the foundations of economic policy, pp. 243–254. New York University Press, New York

Suppes P (1988) Lorenz curves for various processes: A pluralistic approach to equity. Soc Choice Welfare 5: 89–101

Suppes P (1993) The transcendental character of determinism. Midwest Studies Philos XVIII: 242–257

Suppes P (1994) Voluntary motion, biological computation and free will. Midwest Studies Philos XIX: 452–467

Tirole J (1988) The theory of industrial organization. MIT Press, Cambridge

United States Bureau of the Census (1975) Historical statistics of the United States colonial times to 1970, Bicentennial edition, Part II. Washington, D.C.