Operations Research

March–April 1959

BROWNIAN MOTION IN THE STOCK MARKET†

M. F. M. Osborne

U S Naval Research Laboratory, Washington 25, D C

(Received February 6, 1958)

It is shown that common-stock prices, and the value of money can be regarded as an ensemble of decisions in statistical equilibrium, with properties quite analogous to an ensemble of particles in statistical mechanics If $Y=\log_e[P(t+\tau)/P_0(t)]$, where $P(t+\tau)$ and $P_0(t)$ are the price of the same random choice stock at random times $t+\tau$ and t, then the steady state distribution function of Y is $\varphi(Y)=\exp(-Y^2/2\sigma^2\tau)/\sqrt{2\pi\sigma^2\tau}$, which is precisely the probability distribution for a particle in Brownian motion, if σ is the dispersion developed at the end of unit time A similar distribution holds for the value of money, measured approximately by stock-market indices Sufficient, but not necessary conditions to derive this distribution quantitatively are given by the conditions of trading, and the Weber-Fechner law A consequence of the distribution function is that the expectation values for price itself $\mathcal{E}(P)=\int_0^\infty P\,\varphi(Y)(dY/dP)\,dP$ increases, with increasing time interval τ, at a rate of 3 to 5 per cent per year, with increasing fluctuation, or dispersion, of P This secular increase has nothing to do with long-term inflation, or the growth of assets in a capitalistic economy, since the expected reciprocal of price, or number of shares purchasable in the future, per dollar, increases with τ in an identical fashion

IT IS THE PURPOSE of this paper to show that the logarithms of common-stock prices can be regarded as an ensemble of decisions in a statistical steady state, and that this ensemble of logarithms of prices, each varying with the time, has a close analogy with the ensemble of coordinates of a large number of molecules We wish to show that the methods of statistical mechanics, normally applied to the latter problem, may also be applied to the former

Although the results of this paper were first reached inductively from a direct examination of the data on prices, for the sake of clarity we shall present them, at least in part, in a deductive fashion, and compare the

† Read before the U S Naval Research Laboratory Solid State Seminar, February 28, 1958

deductions with observations on prices The fundamental facts, assumptions, and critical points in the derivation of the properties of this ensemble, are given in the numbered paragraphs below These facts and assumptions are *sufficient*, but *not necessary* to obtain agreement of theory with observation

1. Prices move in *discrete* units of $\frac{1}{8}$ of a dollar From this it immediately follows that the natural logarithms of prices also move in discrete units, approximately $1/(8\times\text{price})$ We shall call $\log_e$-units ratio units A ratio unit of $+1\ 00$ corresponds to a ratio of $+2\ 817/1$ We point out for the convenience of the reader, that percentage changes of less than ± 15 per cent, expressed as *fractions* from unity, are very nearly natural logarithms of the same ratio Thus $\log_e[(100+15)/100] \approx +0\ 15$

2. There is a finite, integral number of transactions, or decisions, per unit of time This number may vary from zero to a thousand or more per day for a single stock Hence the decisions also are separated by *discrete* units of time Observationally, the number of decisions per day may be estimated to be not more than the volume in round lots It may be less, for there may be more than one round lot per trade

3. The stimulus of price in dollars, and the subjective sensation of value in the mind of the trader or investor, are related in accordance with the *Weber-Fechner law* As this assumption has engendered some controversy, let us specify precisely its meaning The Weber-Fechner law states that equal ratios of physical stimulus, for example, of sound frequency in vibrations/second, or of light or sound intensity in watts per unit area, correspond to equal intervals of *subjective* sensation, such as pitch, brightness, or noise The value of a subjective sensation, like absolute position in physical space, is not measurable, but changes or differences in sensation are, since by experiment they can be equated, and reproduced, thus fulfilling the criteria of measurability

The Weber-Fechner law is best applicable when there is a single dominant, or primary stimulus Thus in comparing two sounds, or two light intensities, the frequency distribution (pitch or color) must be nearly the same, or the errors of comparison are large Thus, in assuming the Weber-Fechner law, we are relegating earnings, dividends, management, etc , and their future outlook to positions of secondary importance These factors may be important, just as the intermolecular force law is important in molecular problems, in determining departures from the steady state For determining the steady state, we ignore them in the molecular problem, and we ignore their analogs in this problem also

The hypothesis that price and value are related by the Weber-Fechner law can be reached inductively from the raw data by the following rather simple-minded argument Let us imagine that a statistician, trained perhaps in astronomy and totally unfamiliar with finance, is handed a page of the *Wall Street Journal* containing the N Y Stock Exchange

(NYSE) transaction for a given day. He is told that these data constitute a sample of approximately 1000 from some unknown population, together with some of their more important attributes or variables, eleven in all The fact that these eleven were the most important, out of a much larger number obtainable, from annual reports, for example, might be inferred from the fact that this choice of eleven was published every day Our statistician is asked to investigate this population, to determine if it is a homogeneous sample, and what relations (in the probability sense) exist between the variables or attributes listed for each member

The methods of attacking the raw data on such a problem are well-known, especially to biologists, we quote an astronomical reference[1] primarily because of personal familiarity A common first step is the determination of distribution functions Casual inspection of the data reveals that of eleven attributes or variables listed for each member of the population, six, exclusive of the change, are devoted to something called 'price,' evidently a dominant variable even among those so important

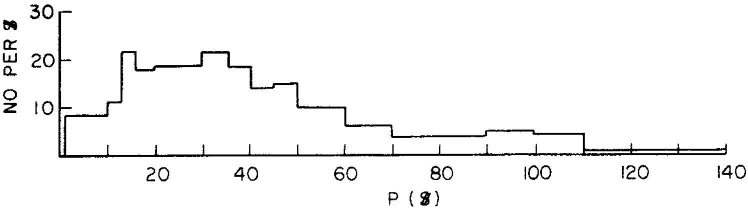

Fig 1 Distribution function of closing prices for July 31, 1956 (all items, NYSE)

as to be published every day On learning that 'close' was the most recent data, our statistician would plot the distribution function of closing price alone for the 1000 members of the sample (Fig 1) Inspection of Fig 1 shows that closing prices on that day were certainly not normally distributed, but the shape suggests that logarithms of prices might be—i e , Fig 1 suggests a logarithmic-normal distribution Figure 2 gives the identical data of Fig 1 with $\log_e$ price as independent variable (*see* reference 1, page 9 for numerical methods)

At this point our statistician will make a 'discovery' and answer one of the questions posed to him A subsidiary maximum around $\log_e P \approx 4\ 5$ ($P \approx \$100$) in Fig 2 suggests that the population contains at least two sub-groups, i e , it is not homogeneous Re-examination of the raw data around $P \approx \$100$ reveals an excessive number of our sample with the attribute 'pfd' (preferred), and plotting the distribution function of these only gives Fig 3 The remaining members of the population, the common, or ordinary ones is plotted in Fig 4 This appears normal, and as a rough test, the cumulated distribution is plotted in Fig 6 At this stage our statistician can say that the population, insofar as the distribution with respect to price is concerned, appears to be divided into perhaps three

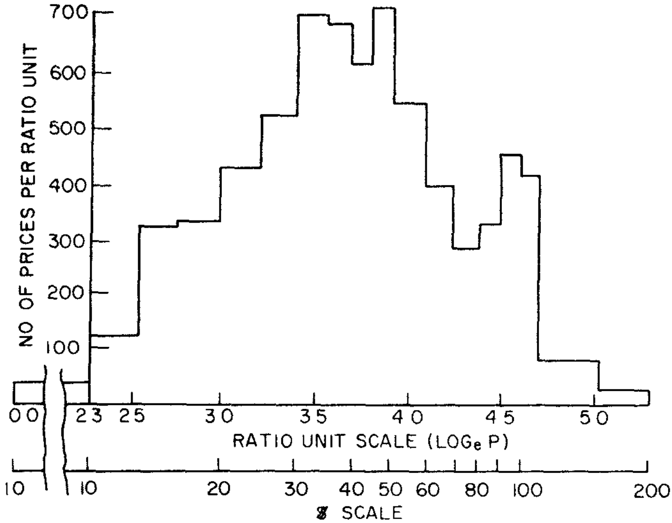

Fig 2 Distribution function for $\log_e P$ on July 31, 1956 (all items NYSE)

classes, if one can regard the two subdivisions of Fig 3 as significant a more searching examination of the data may reveal others (cf Fig 10)

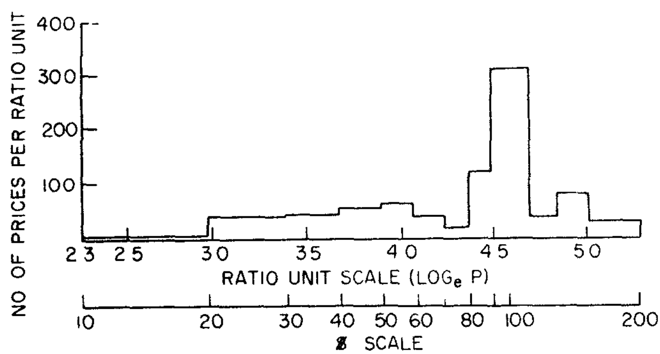

Fig 3. Distribution function of $\log_e P$ for preferred stocks (NYSE, July 31, 1956)

Our statistician likes to choose an independent variable ($\log_e P$ in this case) that renders the data approximately normally distributed The testing of a statistical hypothesis is thereby greatly facilitated, and analogies with many other populations, also normally distributed, may be successfully exploited in trying to understand the new one

A rationale for the use of $\log_e P$ in preference to P as independent variable is also given by the general statistical precept that equal intervals of the argument chosen as independent variable should have equal physical, or in this case psychological, significance, for the data to be most revealing (reference 1, p 6) This choice was confirmed by the resulting 'discovery' of the preferred stocks This 'equal-interval' argument implies that the difference in subjective sensation of profit (or loss), or change in value, for example, between a \$10 and an \$11 price for a given stock, is equal to

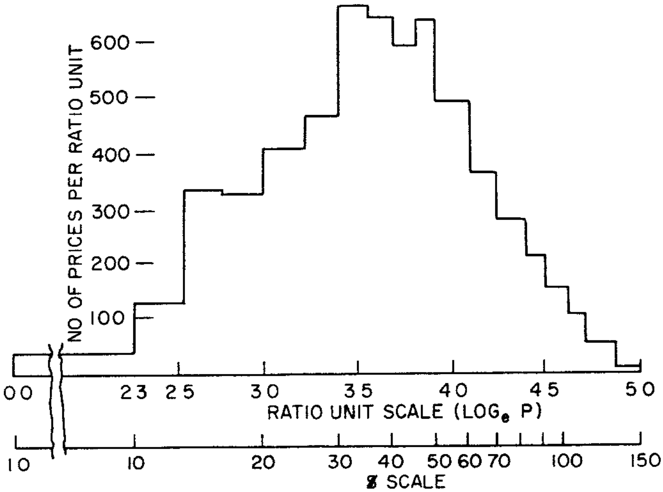

Fig. 4 Distribution function of $\log_e P$ for common stocks (NYSE, July 31, 1956)

that for a change from \$100 to \$110 Thus our statistician is led to hypothecate the Weber-Fechner law, and the dominance of a single variable in the stimulus, from the observational procedure outlined above

Figure 5 gives the closing prices of common stocks on the American Stock Exchange (ASE) and Fig 6 its cumulated distribution, as a rough test of normality

The introduction of the Weber-Fechner law as a working hypothesis now leads our statistician to examine price changes that occur in individual stocks, since by hypothesis the absolute level of price is of no significance, only changes in prices (specifically $\Delta \log_e P$ or the $\log_e$ of price ratios) can be measured by traders or investors Histograms of these for intervals τ of one month are published in *The Exchange* Accumulated distributions for intervals of a month and a year are given in Figs 7 and 8 Note that

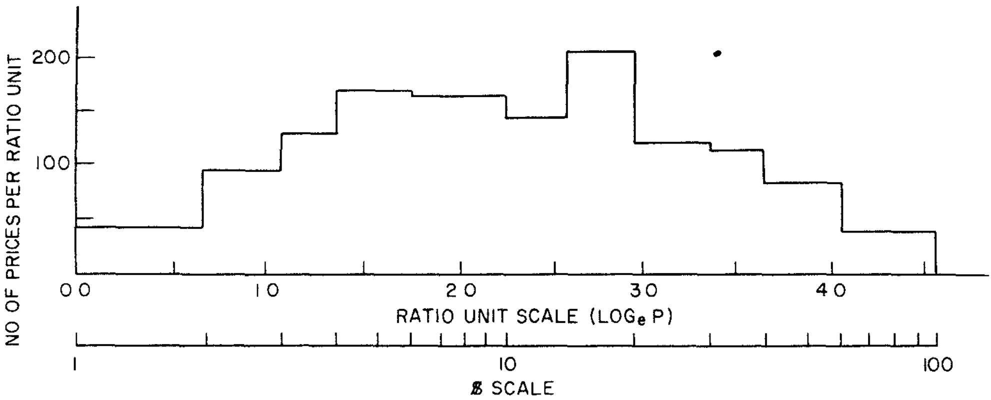

Fig 5 Distribution function of $\log_e P$ for common stocks (ASE, July 31, 1956)

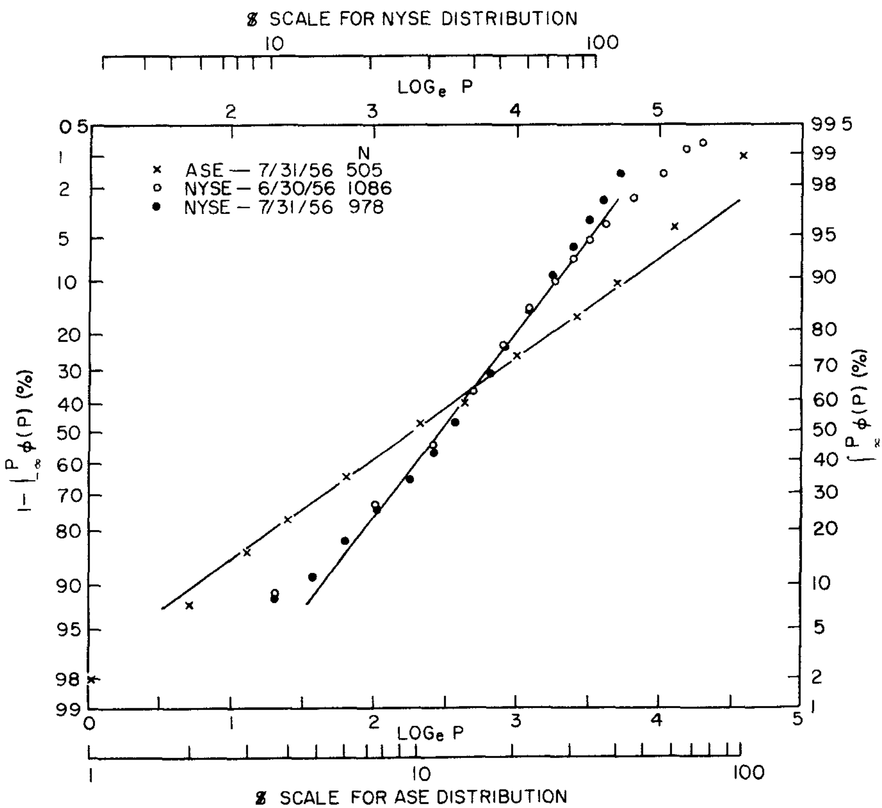

Fig. 6. Cumulated distributions of $\log_e P$ for NYSE and ASE (common stocks)

for both intervals the distributions are nearly normal in ratio units This is slightly less true for percentage units in which the data was originally published The effect is less noticeable in the monthly data, where the percentage changes are small, and hence nearly equivalent to ratio units

This nearly normal distribution in the changes of logarithm of prices suggests that it may be a consequence of many independent random variables contributing to the changes in values (as defined by the Weber-Fechner law) The normal distribution arises in many stochastic proc-

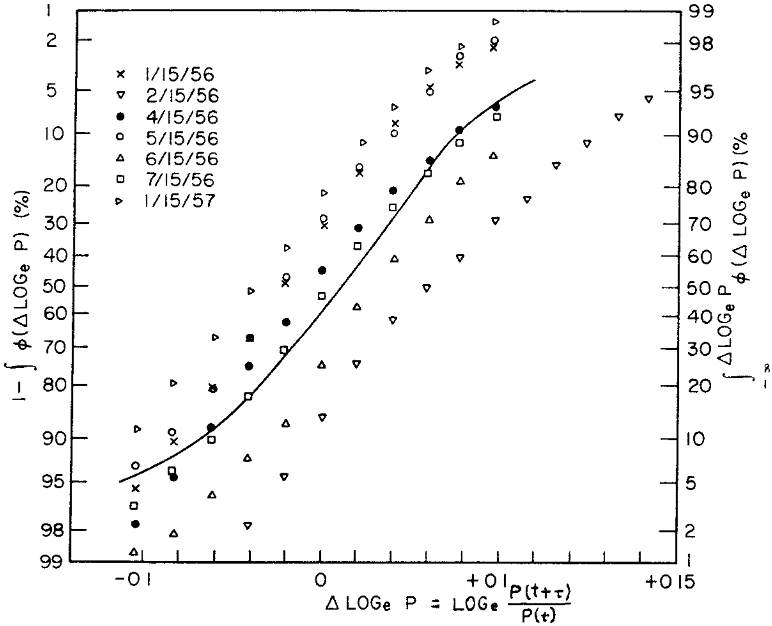

Fig 7 Cumulated distributions of $\Delta\log_e P = \log_e[P(t+\tau)/P(t)]$ for $\tau = 1$ month (NYSE common stocks) These, and also Fig 8, may be regarded as distributions of $S(\tau)$ for fixed $M^*(\tau)$ The solid line is the distribution of $Z(\tau) \approx M(\tau)$, transcribed from Fig 12 for comparison

esses involving large numbers of independent variables, and certainly the market place should fulfill this condition, at least

4. As a fourth element in our analysis, we would like to define a 'logical' decision As an elementary example let us suppose we must make a decision between course of action A, and course of action B We know, or can estimate (in any sense) that course of action A has possible outcomes Y_{A1}, Y_{A2}, with probabilities $\varphi(Y_{A1})$, $\varphi(Y_{A2})$, etc , while a decision for B has possible outcomes Y_{B1}, Y_{B2}, with probabilities $\varphi(Y_{B1})$, $\varphi(Y_{B2})$, etc Then the logical choice is to make a decision for A, or B, for which the expectation value, $\mathcal{E}$, of the outcome, $\mathcal{E}(Y_A) = \Sigma_i\, Y_{Ai}\, \varphi(Y_{Ai})$, or $\mathcal{E}(Y_B) = \Sigma_i\, Y_{Bi}\, \varphi(Y_{Bi})$ is the larger

Evidently decision problems can be much more complicated than this example They may involve several alternatives and sequences of decisions in which the estimate of the probabilities and payoffs (the Y's) are interrelated The general approach is the same to maximize—using

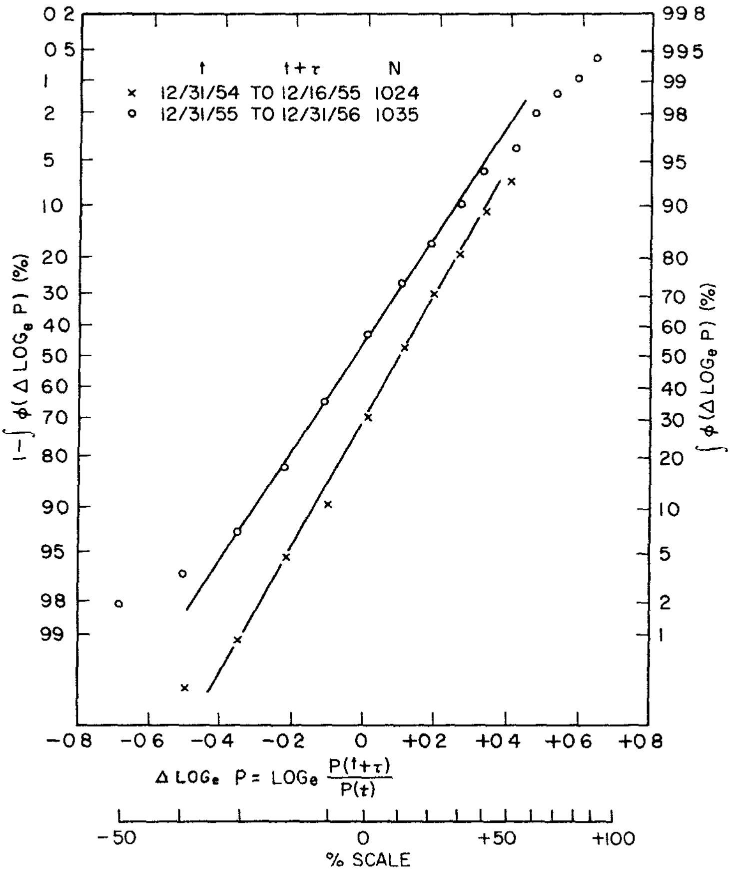

Fig 8 Cumulated distributions of $\Delta\log_e P = \log_e[P(t+\tau)/P(t)]$ for $\tau=1$ year (NYSE common stocks) Data from NYSE Year Book, 1956, and *The Exchange* (February, 1957)

the given information, estimates of probabilities, payoffs, and restraints—the expectation value of the end result

In view of our previous remarks we might illustrate the above example with a stock-market decision A trader has sufficient capital to buy a hundred shares of a corporation, now (time t) selling at $P_0(t)$ He wishes to increase his capital and can choose between A, buying for future sale at some time $t+\tau$, or not buying, B The Y's refer to possible changes in the logarithm of the price of 100 shares, i e , $Y_A(\tau)=\Delta\log_e[100\,P(t)]=\log_e[P(t+\tau)/P_0(t)]$, since by hypothesis it is this quantity that is measurable in the trader's mind There is only one Y_B, zero, the null change with probability one, a certainty The 'logical decision' to buy or not buy is thus determined by whether the estimated expectation value of $Y_A(\tau)$ is positive or negative

We do not claim that the trader sits down and consciously estimates the Y's and $\varphi(Y)$'s, any more than one could claim that a baseball player consciously computes the trajectory of a baseball, and then runs to intercept it The net result, or decision to act, is the same as if they did In both cases the mind acts unconsciously as a storehouse of information and a computer of probabilities, and acts accordingly

Now let us examine the nature of the decisions, of which the published prices gives a numerical measure, concerning the common-stock listings of NYSE These prices represent decisions at which a buyer is willing to acquire stock (and sell money) and a seller is willing to dispose of stock, and hence buy money There are, therefore, in each transaction two types of decisions being made by each participant From what has been said about the anatomy of logical decisions (they need not be consciously logical, but this is the supposition as to how they are reached), we must suppose that for the buyer, his estimate of the expectation value for the change in value ($\Delta\log_e P$) for the stock is positive, while the seller's estimate of the same quantity is negative Presumably, the reverse situation holds in the minds of buyers and sellers for the estimated expectation value for changes in the value of money (their second decision), though we have not yet specified how changes in the value of money are measured in this situation

In view of the equality of opportunity in bidding between buyers and sellers, in accordance with the regulations of the Exchange, it would appear that the most probable condition under which a transaction is consummated, and a price or decision is recorded, is obtained when these two estimates are equal and opposite, or

$$\mathrm{E}\mathcal{E}(\Delta\log_e P)_{\mathrm{S}}+\mathrm{E}\mathcal{E}(\Delta\log_e P)_{\mathrm{B}}=0, \qquad (1)$$

where P denotes price per share, and $\mathrm{E}\mathcal{E}$ the estimate of the expectation value Hence we can say that for the market as a whole, consisting of buyers and sellers,

$$\mathrm{E}\mathcal{E}(\Delta\log_e P)_{\mathrm{M=B+S}}=0 \qquad (2)$$

is the condition under which transactions are most probably recorded A few moments later another transaction may be recorded for the same stock at a slightly different price, and again equation (2) will most probably be applicable, and so on for succeeding transactions One might even argue that in equation (2) the symbol E for estimate could be dropped, since in such buying and selling the decisive estimates are definitive of actual value Or to put it differently, if enough people decide and act on the belief that something is valuable, it is valuable at that time

5. The above contains a critical point in our argument, i e , the most probable condition under which a transaction is recorded is given by equation (2) In words, this states that the contestants are unlikely to trade unless there is equality of opportunity to profit, whether an individual happens at the moment to be a buyer or a seller, of stock, or of money The Exchanges are certainly governed with this end in view, but we also feel that this condition must have obtained even prior to any regulation, since every buyer, once having consummated his trade, now finds himself as a potential seller in the virtual position of his opponent, with whom he was so recently haggling The converse situation applies to the seller, now a potential buyer Under these circumstances it is difficult to see how trading could persist unless prices moved in such a way that equality of opportunity most probably prevailed, and equation (2) expresses this quantitatively, perhaps less as an assumption than as a consequence of assumptions 3 and 4

We now ask, what is the effect of the condition (2) on the distribution function ultimately developed for $\Delta \log_e P$? Our argument follows closely one originally given by GIBBS for an ensemble of molecules in equilibrium The actual distribution function is determined by the conditions of maximum probability (reference 3, p 79)

6. Assuming the decisions for each transaction in the *sequence* of transactions in a single stock are made independently (in the probability sense), then under fairly general conditions outlined below, we can expect that the distribution function for $Y(\tau) = \log_e[P(t+\tau)/P_0(t)]$ will be normal, of zero mean with a dispersion $\sigma_{Y(\tau)}$ which increases as the square root of the number of transactions If these numbers of transactions (the 'volume') are fairly uniformly distributed in time, then $\sigma_{Y(\tau)}$ will increase as the square root of the time interval, i e , $\sigma_{Y(\tau)}$ will be of the form $\sigma\sqrt{\tau}$, where σ is the dispersion at the end of unit time

7. Mathematically we may express this as follows Suppose we have k independent random variables $y(i)$, $i = 1, \quad , k$,

$$y(i) = \Delta_{i\delta} \log_e P = \log_e[P(t+i\delta)/P(t+\{i-1\}\ \delta)],$$

where $P(t)$ is the price of a single stock at time t and δ is the small time interval between trades † Assume that each $y(i)$ has the same dispersion $\sigma(i) = \sigma'$, then after k trades, a time $\tau = k\delta$ later, we define $Y(\tau)$ as

$$Y(\tau) = Y(k\delta) = \sum_{i=1}^{i=k} y(i) = \log_e[P(t+\tau)/P(t)] = \Delta_\tau \log_e P(t) \qquad (3)$$

† Indices i, j in parentheses will refer to independent random variables in a sequence in time As subscripts, i, j will refer to independent variables at the same time (different stocks)

We also have the dispersion of Y

$$\begin{aligned}\sigma_{Y(\tau)} &= \sqrt{\mathcal{E}(Y^2) - [\mathcal{E}(Y)]^2} \\ &= \sqrt{\textstyle\sum_{i=1}^{i=k} \sigma^2(i)} = \sqrt{k}\,\sigma' = \sqrt{\tau/\delta}\,\sigma'\end{aligned} \tag{4}$$

Hence $\sigma = \sigma'/\sqrt{\delta}$ is the dispersion developed at the end of unit time, i.e., in a number $1/\delta$ of trades.

The central limit theorem[5] assures that $Y(\tau)$ will approach a normal distribution for large k whatever the distribution function of the $y(i)$.

The above considerations have an obvious analog in the diffusion of a molecule undergoing collisions with its neighbors. Regardless of the intermolecular force law (cf. the factors ignored in hypothecating the Weber-Fechner law), the dispersion in the probability distribution of position of a particle initially (at time t) located at some point will increase as the square root of the time interval τ after t. The phenomenon of the persistence of velocities,[2] which the stock market also possesses[7] does not alter this conclusion.

THE OBSERVATIONAL DATA

Let us now examine the data to see in what particulars the above expectations are fulfilled. Figures 7 and 8 support, at least approximately, the conclusion of normality for $Y(\tau)$, at least for intervals $\tau = 1$ month and 1 year. We would now like to estimate how the distribution function of $Y(\tau)$ changes with the time interval τ. Owing to the wealth of data, and for purposes of computational simplicity, we shall not evaluate $\sigma_{Y(\tau)}$ directly. Instead, we shall evaluate the semi-interquartile range (s.q.r.) of $Y(\tau)$, or one-half of the range (in ratio units) between the 25 per cent and 75 per cent points of a sample. For a normally distributed population, which Figs. 7 and 8 indicate is quite closely the case here, s.q.r. = 'probable error' of $Y(\tau) = 0.6745\ \sigma_{Y(\tau)}$. In any event the s.q.r. has a definite statistical interpretation, whether the distribution is normal or not.

To evaluate the s.q.r. we take a random sample of common stocks at some random starting date, t, write down the prices of each on that date and the prices of the same stocks at intervals τ of a day, week, month, two months, etc., later. All of these prices are then divided by their corresponding starting (t) prices, and the ratios of each plotted for the various intervals. One-half of the interquartile range is obtained by taking one-half of the natural logarithm of the ratio of ratios from the 25 per cent to 75 per cent point of the sample. This s.q.r. is then plotted in Fig. 9. Similar data on a *trading-time* scale, appears in Fig. 10. As can be seen from an inspection of these two figures for various starting dates from

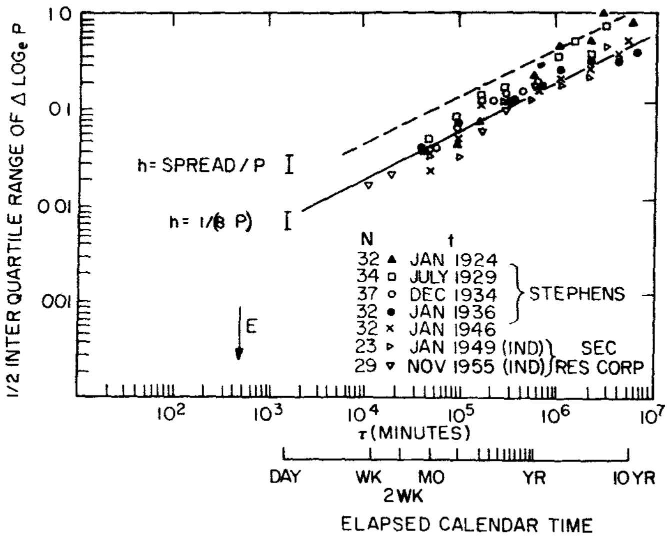

Fig 9 Semi-interquartile range of $\Delta\log_e P = \log_e[P(t+\tau)/P(t)]$ for τ as a calendar time interval (NYSE common stocks) N = number of stocks in sample Data from F W Stephens and the Securities Research Corporation This data may be regarded as 0 6745 $\sigma_{S(\tau)}$ for an assumed normal distribution The straight lines are transcribed from Fig 11

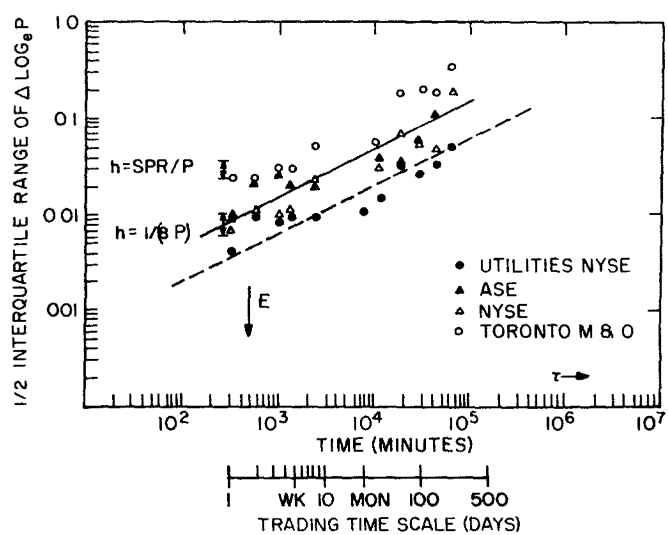

Fig. 10 Semi-interquartile range of $\Delta\log_e P = \log_e[P(t+\tau)/P(t)]$ for τ as a trading time interval, for common stocks (all data from the *Wall Street Journal*) Samples from NYSE were taken under letter H, from ASE under letter M, from Toronton under mines and oils The utilities were the first 20 listed by Securities Research Corporation Starting time t, September 21, 1956 N=20 for all samples This data may be regarded as 0 6745 $\sigma_{S(\tau)}$ for an assumed normal distribution

1924 to 1956, and intervals of one day to 12 years, the s q r does indeed increase with the square root of the time, corresponding to a slope of $\frac{1}{2}$ Even the data beginning in July 1929 is not exceptional in this regard Continuity of the data from the two figures may be obtained by noting the shift from a trading- to calendar-time scale, which occurs mostly between a day and a week Note that this measure of the dispersion of Y, 0 6745 $\sigma_{Y(\tau)}$ from Figs 9 and 10, is from a *sample* of N for some single interval τ

The years 1924–1956 were the limits of readily accessible data, published by F W Stephens and the Securities Research Corp The random sampling was achieved by paging through the publications and flipping a coin to decide which page and common stock, of the NYSE, to select The short range data of Fig 10 were taken directly from the *Wall Street Journal* in a similar fashion

In order to clarify the next step, let us anticipate the results of an inductive analysis of the data by describing some models which have many of the features of our market—the ensemble of 1000 or more logarithms of individual stock prices, as functions of the time

MODEL I Let us first imagine 2000 pennies grouped in 1000 pairs All 1000 pairs are tossed simultaneously at intervals δ, or $1/\delta$ tosses per unit time Heads count $+1$, tails -1, and we record the payoff 2, 0, or -2, $y_j(i)$ of the jth pair ($j = 1 \quad 1000$) of the ith toss We also record the mean of a *sample* N in number for each toss, $m(i) = 1/N \sum_{j=1}^{j=N} y_j(i)$, evidently this will be very close to zero for large N We also record the cumulated sum after k tosses or after an interval τ of each pair, and of the mean, i e, $Y(\tau)$ and $M(\tau)$ where $\tau = k\delta$ We also ascribe an arbitrary starting point for each random walk $Y(\tau)$, which, to preserve the analogy with our previous notation, we shall call $\log_e P_j(t)$ for the jth pair We deliberately add and subtract this arbitrary constant, to emphasize that $Y_j(\tau)$ is the deviation from some arbitrary starting point (c f Weber-Fechner law)

$$Y_j(\tau) = [\log_e P_j(t) + \textstyle\sum_{i=1}^{i=k} y_j(i)] - \log_e P_j(t), \tag{5}$$

$$\begin{aligned} M(\tau) = \textstyle\sum_{i=1}^{i=k} m(i) &= (1/N) \textstyle\sum_{j=1}^{j=N} \sum_{i=1}^{i=k} y_j(i) \\ &= (1/N) \textstyle\sum_{j=1}^{j=N} Y_j(\tau) \end{aligned} \tag{6}$$

Evidently the dispersion of the Y's, $\sigma_{Y(\tau)}$ can be computed theoretically, and also estimated from the data, among other methods, by a method similar to that above on stock prices The dispersion of $M(\tau)$, $\sigma_{M(\tau)}$, since $M(\tau)$ is a single random walk, can also be computed theoretically, but must be experimentally obtained from samples of nonoverlapping intervals of duration τ

With the above model, it is not difficult to see that $\sigma_{M(\tau)} \approx \sigma_{Y(\tau)}/N^{1/2}$ and both $\sigma_{M(\tau)}$ and $\sigma_{Y(\tau)}$ increase with the square root of k, or the square root of the time interval τ if the tossing rate is constant In this model $M(\tau)$ and $Y(\tau)$ will both be normally distributed, for large τ or equivalently large k

MODEL II Let us now consider a second ensemble consisting of 1000 pennies and one gold piece The payoffs of heads and tails for each coin are the same, as before, the payoffs for gold and copper coins may or may not be the same We also form 1000 pairs from this ensemble, each pair having the single gold piece in common, and determine as before $Y_j(\tau)$ and $M(\tau)$ The *individual* random walks $Y(\tau)$ are statistically identical with those of Model I The difference in the two *ensembles* shows up in $m(i)$ and in $M(\tau)$ For large N, $m(i)$ will unambiguously reveal whether the gold coin came up heads or tails, and the dispersion of $M(\tau)$ for large N will be almost identical with that determined for the $Y(\tau)$'s, not $1/(N)^{1/2}$ smaller, *if* $\sigma_{Y(\tau)}$ *is determined from a sample* for *one* interval, as in Figs 7–10 $M(\tau)$ will, for large N, be almost the same as the random walk generated by the gold coin alone As before, both $\sigma_{Y(\tau)}$ and $\sigma_{M(\tau)}$ will increase as $\tau^{1/2}$ and $Y(\tau)$, $M(\tau)$ will be normally distributed If we should modify the payoff of the gold coin to a number other than that for the copper coins, this will show experimentally in $\sigma_{M(\tau)}$

MODEL III In this model we modify Model II by adding a few copper coins, but still form 1000 groups (no longer all pairs—some comprise three or more coins) Every group has at least one, the gold coin, in common We further modify slightly the payoff of a few of the coins, increasing the amount of head and tail values in some, decreasing it in others, but still keeping them equal but opposite in sign The payoff of the gold coin may be increased for some groups, decreased for others, corresponding to 'leverage' and 'defensive' stocks We again form the $y_i(j)$, $m(j)$, $Y_i(\tau)$, $M(\tau)$ of these 1000 groups as before For this case the distribution of the $Y(\tau)$'s determined from a sample, as in Figs 7–10, will not be quite normal since the component dispersions are not identical However, $\sigma_{Y(\tau)}$ or the interquartile range of $Y(\tau)$ will increase with $\tau^{1/2}$ as before, and the same statement will also be true for $M(\tau)$ $M(\tau)$ and $m(i)$ still will, for large N, reveal rather reliably the behavior of the gold coin alone

Our problem is the inverse of deriving theoretically the properties of these models We have the 1000 or so random walks or stock prices, and hence can form numerical values for $Y_j(\tau) = \log_e[P_j(t + \tau)/P_j(t)]$ We do not know how the Y's were generated, and we wish to examine them to find out which model, if any, might represent best their behavior

8. At this point we put forth the hypothesis, and we believe the data will support it, that Model II will represent the behavior of our ensemble of stock-market prices The data will require small, but nevertheless detectable modifi-

cations in the direction of Model III The payoffs are in ratio units, not dollar changes, and can be estimated from the discrete nature of the process described in 1 and 2

9. We further assert, exploiting the analogy of Model II with the elementary-decision process described for the market, that $M(\tau)$ will estimate changes in the value of money in the minds of buyers and sellers The value of money is not measurable by the Weber-Fechner hypothesis, but changes in the value of money should be

An estimate of the expectation value for changes in the value of money was an element common to every decision, in precisely the same sense that the gold coin payoff was common to every pair of Model II Approximately, but not exactly, this change in the value of money is given by changes in the ordinary stock-market indices We should however, note the restriction to buyers and sellers of common stocks We do not know whether $M(\tau)$ is the same for those who deal in bonds, preferred stocks, commodities, or wherever else money may go

Let us now return to the data For our purpose the ideal index of stock market prices, as a function of the number of transactions from some arbitrary starting date would be

$$\overline{\log P(k)} = (1/N)\sum_{i=1}^{i=N} \log P_i(k)$$

formed from N individual stock prices However, if the number of transactions per day does not fluctuate too violently it would be sufficient to use, without appreciable error

$$\overline{\log P(t)} = (1/N)\sum_{i=1}^{i=N} \log P_i(t)$$

or the equivalent geometric mean

$$\bar{P}(t)_{\text{geom}} = \left[\prod_{i=1}^{i=N} P_i(t)\right]^{1/N} \tag{7}$$

Stock market indices are not computed this way either However, there are numerous arithmetic averages of the type

$$\bar{P}(t)_{\text{arith}} = (1/N)\sum_{i=1}^{i=N} P_i(t) \tag{8}$$

We hope to show that changes in the logarithm of $\bar{P}(t)_{\text{arith}}$ are approximately the same as changes in $\overline{\log P(t)}$ and that the dispersion of the changes thus computed can be corrected, approximately, to the dispersion of $M(\tau) = \Delta\overline{\log P(t)} = \log[\bar{P}_{\text{geom}}(t+\tau)/\bar{P}_{\text{geom}}(t)]$

The most comprehensive series of indices of the type (8) were computed by Cowles [7] The number of stocks, N, varied from 8 to 18 for the years 1831–1936, $N=15$ since 1897 Cowles, moreover, computed not the dispersion of the changes of his index, but the absolute value of percentage changes for nonoverlapping intervals of τ from 20 minutes to 12 years In other words, Cowles computed

$$Z(t_i, \tau) = \sum_{j=1}^{j=N} [P_j(t_i+\tau) - P_j(t_i)] / \sum_{j=1}^{j=N} P_j(t_i), \qquad (9)$$

and published values of $\overline{|Z(\tau)|} = (1/L) \sum_{i=1}^{i=L} |Z(t_i, \tau)|$ Here L is the number of nonoverlapping intervals used, from 1831–1936 or less From our discussion under Model I of how to compute the dispersion of $M(\tau)$, and since percentages and ratio units are closely related, this $\overline{|Z(\tau)|}$ is almost, but not quite, what we seek

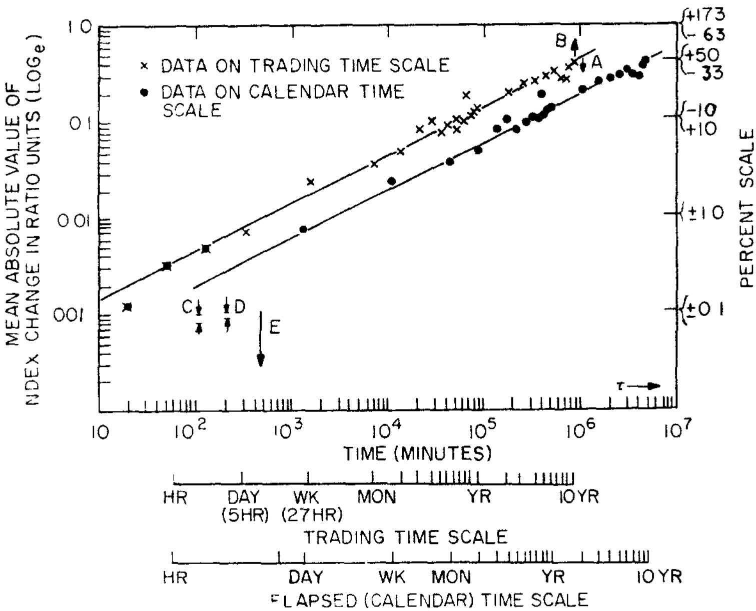

Fig 11 Mean absolute value of index changes according to Cowles, in ratio units, as a function of time interval The arrows A–D denote approximate corrections to convert the data to s q r $= 0\ 6745\ \sigma_{M(\tau)}$ of a normal distribution

In Fig 11 we have converted Cowles's value of $\overline{|Z(\tau)|}$ to ratio units (assuming all percentages were positive) and plotted them In order to compare this data with that of Figs 9 and 10, we have computed the small corrections A, B, C, D outlined below to convert this $\overline{|Z(\tau)|}$ to the semiquartile range of the variable $M(\tau)$, assumed normally distributed While these corrections are uncertain, they are small and some of them tend to cancel out

Let us first compare $Z(\tau)$ with $M(\tau)$

$$M(\tau) = (1/N) \sum_{j=1}^{j=N} \log_e [P_j(t+\tau)/P_j(t)] = (1/N) \sum_{j=1}^{j=N} Y_j(\tau) \quad (10)$$

In terms of Y's, Cowles's function becomes, since $P(t+\tau)=P(t)\,e^Y$,

$$\begin{aligned} Z(\tau) &= \sum_{j=1}^{j=N} P_j(t)\{\exp[Y_j(\tau)]-1\} \Big/ \sum_{j=1}^{j=N} P_j(t) \\ &= \sum_{j=1}^{j=N} P_j(t)[Y_j(\tau)+\tfrac{1}{2}\,Y_j^2(\tau)+\quad] \Big/ \sum_{j=1}^{j=N} P_j(t) \qquad (11) \\ &= \sum_{j=1}^{j=N} P_j(t)\;Y_j(\tau) \Big/ \sum_{j=1}^{j=N} P_j(t)+O[Y^2(\tau)] \end{aligned}$$

Thus we see that, neglecting terms of order Y^2, $M(\tau)$ is an unweighted mean of the Y_j's, whereas $Z(\tau)$ is weighted in favor of the higher-priced stock at the start of the interval, t The systematic effect of this price weighting (which was, to be sure, changed from time to time), we do not know, but can only observe that over the entire history of the Exchange, which Cowles sampled, the same stocks undoubtedly became high- and low-priced several times If we can justifiably make the assumption that over the entire history of the Exchange, high-priced stocks do not behave systematically differently from a random sample, and can neglect terms in Y^2, then indeed $Z(\tau)\approx M(\tau)$ The mean absolute value of the Z's, which Cowles computed, is closely related to the dispersion It is nearly the so-called 'mean error '

Cowles's data have been corrected to ratio units (assuming all percentages were positive) and plotted in Fig 11 His data follow a slope ½, or square-root-of-time diffusion law very nicely indeed The data have been plotted twice, once on a trading-time scale (as Cowles gives it) and once on a calendar time scale Evidently the trading-time scale is the significant one, since the data are more nearly continuous on that basis, whereas on the basis of calendar time, there is a shift in the two plots, mostly between intervals of less than a day to a week, corresponding to the difference between a 5- to 24-hour day, and a 5- (or 5½) to 7-day week

For comparison, the solid lines of Figs 9 and 10 are transcribed from Fig 11 for the appropriate time scale It is not at all obvious that these lines should agree with the data as well as they do, since Figs 9, 10, and 11 refer to apparently quite different measurements Figures 9 and 10 give the relative diffusion of stocks with respect to their median, or 'center of gravity,' while Fig 11 represents the increasing fluctuation, with time interval, of the center of gravity itself This agreement is precisely what one would expect, however, if stock price changes were generated by Model II

This agreement in the diffusion rates of $Y(\tau)$ and $M(\tau)\approx Z(\tau)$ suggests that the distribution functions of Y and Z should be very similar Examples of the distribution function of Y (*from a sample for a fixed interval*) are given in Figs 7 and 8 As a typical example of an arithmetic average, we have taken the Dow industrial index and evaluated the distribution function of its monthly changes (Z) in ratio units from 1925 to 1956 The cumulated distribution is given in Fig 12, which is also drawn on Fig 7, for comparison Evidently the distributions are quite comparable Both show approximately the same dispersion, both are nearly normal with a slight excess corresponding to the smaller slope at 5 per cent and 95 per cent extremes A part of this excess in Fig 12 may be due to the linear weighting, and the neglected $O(Y^2)$ terms in comparing Z and M

If we admit the approximate assumption, from Fig 12, that $Z \approx M$ is distributed normally, then we can compute the corrections denoted by the arrows A, B, C, D, in Fig 11, which would convert Cowles's absolute percentage changes to a semiquartile range, or 'probable error' for M

If we have a variable Z, normally distributed about a mean $Z_0 \neq 0$, and if, as is the case here, Z_0 is small compared to the dispersion of the distribution, then approximately

$$\overline{|Z-Z_0|} \cong \overline{|Z|} - Z_0 \int_0^{Z_0} \varphi(Z)\, dZ \tag{12}$$

In other words, to convert Cowles's deviation from zero (Fig 11) to an absolute deviation from the mean, we subtract off the shift from zero to mean, Z_0, times the fraction of the population lying in this range This turns out to be negligible for all intervals less than a year However, as an extreme example, let us suppose both Z_0 and $\overline{|Z|} = +0\ 4$ in ratio units, corresponding to a 50 per cent secular advance in the average over a ten-year period, a figure also of the order of the s q r of Y (one-fourth of the population) for this interval (*see* Figs 9 and 11) In this case $\overline{|Z-Z_0|} \approx 0\ 4-(\tfrac{1}{4})\ 0\ 4 = 0\ 3$ The reduction corresponding to this hypothetical case corresponds to arrow A of Fig 11

The above correction is to some extent compensated by the conversion of Cowles's percentages to ratio units Positive and negative percentages and ratio units are nearly equivalent when small, but not when large, corresponding to the larger time intervals The data in Fig 11 assumed that all percentages were positive The opposite extreme, assuming all were negative, would have little effect except for the longer time intervals As an extreme case, the ten-year point (+51 64 per cent) corresponds to 0 416 ratio units, −51 64 per cent would correspond to 0 708 ratio units, as indicated by arrow B in Fig 11 This is one effect of the $O(Y^2)$ term neglected in (11)

Under the assumption of a normal distribution, the absolute deviation from the mean is $\sqrt{2/\pi}/0\ 6745$ larger than the s q r, or 'probable error' The displacement C on Fig 11 corresponds to this reduction, applicable to all of Cowles's data

Finally, we have to estimate the effect of the linear weighting with respect to price, noted in equation (11) The analysis necessary to determine the relation of the dispersion or s q r of Z to that of M unambiguously rejects Model I in favor of Model II

Starting with equation (6) let us first assume that the $Y_j(\tau)$ are independent with respect to j Then we have

$$\sigma^2_{M(\tau)} = \mathcal{E}[M^2(\tau)] - \{\mathcal{E}[M(\tau)]\}^2 = (1/N^2)\sum_{j=1}^{j=N} \sigma^2_{Y_j(\tau)} \tag{13}$$

The assumption of the independence of the $Y_j(\tau)$ is not the same as that previously mentioned, that sequences of decision on the same stock are independent, but assumes also that simultaneous decisions on different stocks are independent This corresponds to Model I, for which each $Y_j(\tau)$ (or pair of coins) has *no* random variable in common with any other

If, moreover, all the $Y(\tau)$ have the same variance $\sigma^2_{Y(\tau)}$, then

$$\sigma_{M(\tau)} = \sigma_{Y(\tau)}/\sqrt{N} \tag{14}$$

Under the same assumptions and neglecting $O(Y^2)$ in (11),

$$\begin{aligned} \sigma^2_{Z(\tau)} &= \sigma^2_{Y(\tau)} \sum_{j=1}^{j=N} P_j^2(t) / \left[\sum_{j=1}^{j=N} P_j(t)\right]^2, \\ \sigma_{Z(\tau)} &= (\sigma_{Y(\tau)}/\sqrt{N})[\overline{P^2(t)}/\overline{P(t)}^2]^{1/2}, \end{aligned} \tag{15}$$

where the bar now refers to arithmetic averages of sample prices at a given time t The factor $[\overline{P^2(t)}/\overline{P(t)}^2]^{1/2}$ we estimate from Fig 1 less the preferred stock, as a typical example, to be 1 16, corresponding to displacement D This small reduction applies to Cowles's data for all time intervals

In deriving $\sigma_{M(\tau)}$, $\sigma_{Z(\tau)}$, and their approximate equivalence, the assumption preceding equation (13), a property of Model I, is most emphatically rejected $\sigma_{Y(\tau)}$ can be evaluated from the s q r in Figs 9 and 10 The above equations (14) and (15) indicate that $\sigma_{Z(\tau)}$ or $\sigma_{M(\tau)}$ are both $\approx \sigma_{Y(\tau)}/\sqrt{N}$, where N is the number of stocks in the sample, say 15 as an average figure for Cowles's data This would reduce $\sigma_{Z(\tau)} \approx \sigma_{M(\tau)}$ to (1/3 9) $\sigma_{Y(\tau)}$, which as Figs 9 and 10 show is most definitely not the case $\sigma_{M(\tau)}$ and $\sigma_{Y(\tau)}$ are in fact very nearly equal Arrow E shows the expected reduction in Cowles's data, because of their $\sqrt{15}$ factor

Thus an assumed sampling of one variable, $y_i(j)$, as in Model I (*one* in the sense that all the copper coin pairs were supposed to generate statistically identical random variables), fails to represent the data Let us next assume we are sampling two variables One variable, $s_j(i)$, we assume varies, corresponding to the copper coins, independently in time and from one stock (or coin) to the next, the second, $m^*(i)$, corresponding to the gold coin, is common to all stocks (or pairs of coins) at a given time, t, but may vary independently from one value of t to another So we have

$y_j(i)=m^*(i)+s_j(i)$, and form, as before the sums, or random walks, $Y_j(\tau)=M^*(\tau)+S_j(\tau)$, $j=1$–1000 We carry through the above derivation with assumed independence of the s_j and that m^* is independent of s_j We are now trying to fit our data to Model II Evidently $M(\tau)$, a mean of a sample of size N of $Y_j(\tau)$ will *estimate* $M^*(\tau)$ accurately, and we find, in fact

$$\sigma^2_{M(\tau)}=\sigma^2_{M^*(\tau)}+\sigma^2_{S(\tau)}/N \qquad (16)$$

The N now makes the second terms of equations (15) and (17) quite negligible by comparison to the first term The adequacy of Cowles's data on Z in representing the market variable M^*, or its estimate, the sample mean, M [from equations (10) and (11)], within the errors denoted by A to D is still maintained

At this point we should like to distinguish between the dispersion of $\sigma_{Y(\tau)}$ as defined and determined by the data for a sample of Y_j's of size N for a single time interval τ (as in Figs 7–10) or a dispersion for Y determined by taking a single $Y_j(t)$ and determining this dispersion from N nonoverlapping intervals of length τ (the method we used for determining $\sigma_{M(\tau)}$) So far as Model I is concerned, either method would give the same result, but this is not the case for Model II The first method gives for Model II, $\sigma_{S(\tau)}$, and it is so identified on Figs 9 and 10 The second method would give $\sqrt{\sigma^2_{S(\tau)}+\sigma^2_{M^*(\tau)}}$ as in (21) Model II is a nonergodic ensemble, that is, a sample of N members over a single interval τ, has properties different from that given by one member sampled for N separate intervals of length τ

This does not yet explain why $\sigma_{S(\tau)}$, now identified as being determined by Figs 9 and 10, happens to be so nearly equal to $\sigma_{M(\tau)}\approx\sigma_{M^*(\tau)}$ (Fig 11 with corrections) Returning to our Model II, we had no way of knowing in advance that the allowed payoffs of the copper and gold coins should be the same, but the data, within the limits of rather uncertain errors, tells us that this is indeed the case

It should now be plausible why we identify the market variable, or $-M^*(\tau)$, if we wish to preserve conventional ideas on sign, with changes in the value of money, corresponding to the variable generated by our gold coin As we saw from our discussion of the recorded decisions, an estimate of the expectation value for changes in the value of money was an element common to every decision (of opposite sign to that concerning stock), in exactly the same sense that the payoff of the gold coin was an element common to every random variable of Model II It is also exceedingly plausible, as the data indicate is the case, that $M(\tau)$ and $S(\tau)$ should have the same distribution function and diffuse in the same way To put it another way, the NYSE is a market for money in exactly the same sense that it is for the securities of any given corporation. Certainly for the era covered by Cowles's data, a dollar represented a share in the assets of Fort Knox in exactly the same sense that a stock certificate of General Motors represented a share in the assets of that corporation. Under condition of trading in statistical equilibrium, why should these changes in value not diffuse in the same way?

The above analysis of the variance of $Y(\tau) = \log_e[P(t+\tau)/P_0(t)]$ about its median or mean led to the discovery of two variables S and M^* being sampled in our data, and that $Y_j(\tau)$ should be expressed as $S_j(\tau)+M^*(\tau)$, which for simplicity we have approximated as independent, as in Model II. One might reasonably ask whether or not a closer examination of the data might not reveal still other variables. That is to say, should not Model II be modified in the direction of Model III? The answer to this question is undoubtedly 'yes.' One could hardly claim that the data on 1000 or more random time series could significantly support not more than two significantly different variables. However, the determination of just how many and which ones, or combination of them, are significantly independent, would be an exceedingly tedious task. Such an analysis could be performed and would undoubtedly confirm much that is common knowledge. We give a very simple example below.

Figure 10 gives the s.q.r. for several groups of stocks which, on the basis of common knowledge, might be expected to behave in a significantly different way. Evidently the utilities diffuse at a significantly smaller rate than a random sample from ASE or NYSE, which in turn diffuse less rapidly than the mines and oils from the Toronto Exchange. All, however, follow approximately a $\sqrt{\tau}$ diffusion law corresponding to a slope of ½. Note that the data for NYSE extends the range of Fig. 9 down to one day, when the shift from trading time to calendar time is taken into account. The data from NYSE and ASE are not significantly different, to the author a rather surprising result.

Aside from the direct evidence of Fig. 10, close examination of the data of Figs. 7 and 8 gives indirect evidence for the presence of more than two distinguishable variables in the data. The plots of Figs. 7, 8, and 12 are not precisely straight, there is a systematic flattening of the curves at the upper and lower ends corresponding to a larger 'tangential dispersion' in the data at these limits. The slope corresponds to the 'nearest fitting' dispersion in such a plot. Now, if we are sampling a population consisting of several different components, all with the same mean (cf. the gold coin) those components with the larger dispersion will always dominate the data at the extremes of the distribution function, which is precisely what the above effect shows. To put the argument another way, if one could imagine a stock exchange in which the only equities traded were public utilities, and Canadian mines and oils, in any mix, then the distribution function of $\Delta \log_e P$, similar to Figs 7 and 8, would in its extremes be dominated by the latter, and show to an exaggerated degree the flattening just noticeable in Figs 7 and 8 This shows that the assumption of the equality of the σ's preceding equation (3) and implied for Model II, cannot be rigorously correct

The flattening in Fig 12 may be due in part to the 'mathematical errors' introduced by the use of arithmetic averages and the different dis-

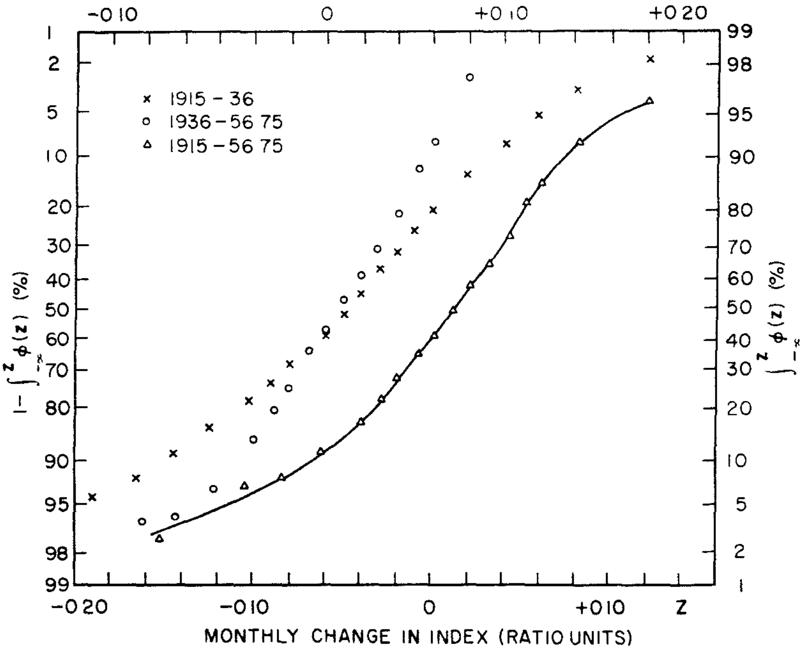

Fig 12 Cumulated distribution function of the monthly changes $Z(\tau)$ in the Dow Industrial Index, 1915 to 1956 75, in ratio units The top scale is for the separate intervals 1915–1936 and 1936–1956 75

persions in its components To the extent that there is a contribution to the flattening over and above these, it would reflect a change in the payoff of the gold coin, or correspondingly in the dispersion or diffusion rate of money The data covers the period 1916–56, during which time there have been some changes made in the definition of a dollar Hence, such an effect should not be surprising, and it would be interesting to try to pinpoint the dates of change by careful analysis of short-range data The two intervals noted on Fig 12 have, in fact, slightly different slopes

Our interpretation of the data on $\log_e P$ in terms of Model II led us to suppose proportionality between τ, the time interval between observations, and k, the corresponding index denoting the number of elapsed transactions, i e, $\tau = k\delta$ While this may be approximately true 'on the average' it is certainly not true in detail The 'volume' does vary from one day to the next, hence presumably also the number of transactions, unless we make the most unlikely assumption that the number of round lots per trade varies in just such a way as to keep the transaction rate constant It is a matter of common observation that the volume tends to be larger when the market as a whole (i e, all stock prices), heaves up or down most rapidly, or in terms of our model, the volume and hence transaction rate, tend to be larger when absolute values, $|M(\tau)|$, are large Now, if the transaction rate is large, so also will be the diffusion rate, hence this effect should show up in our data if we were to plot the semi-interquartile

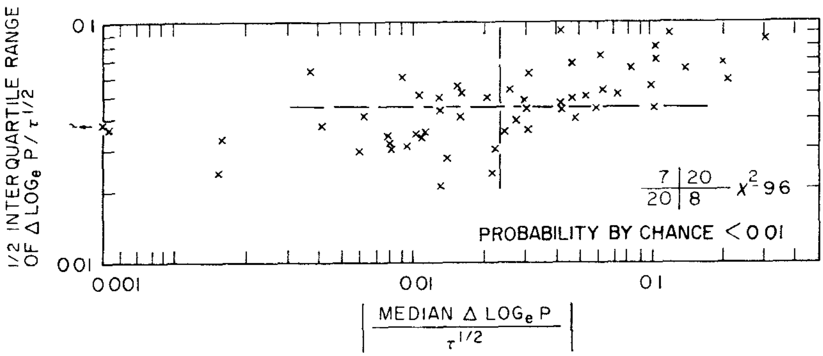

Fig 13 Test for the dependence of $\sigma_{S(\tau)}$ on $|M|$ The semi-interquartile range of $\Delta\log_e P$ is 0 6745 $\sigma_{S(\tau)}$, and the median of $\Delta\log_e P$ is $\approx M$ τ is measured in months

range of $\Delta\log_e P$ (from Figs 7 and 8) against the corresponding median of $\Delta\log_e P$ from the same data Since both of these are already known to increase with $\tau^{1/2}$, we have divided them by this factor, $\tau^{1/2}$, so that data over intervals of all length will be comparable This is done in Fig 13 which shows a small but significant dependence in the sense expected, i e, when the reduced semi-interquartile range is large, so also is the reduced absolute displacement of the market as a whole The numbers at the corner of Fig 13 refer to a chi-square test for significance of the dependence from a 2 × 2 contingency table about the medians of the graph [5] This effect accounts, at least in part, for the scatter of the points in Figs 9 and 10 around a straight line of slope ½

Our discussion of the equality of the estimated expectation values for buyers and sellers, for $Y = \Delta\log_e P$ led us to believe that the dispersion for

Y should increase as the square root of the time interval This was amply confirmed by the data Closer examination of Cowles's data, which approximated the dispersion of the mean of a sample of Y, suggested that Y should be expressed as the sum of two variables nearly independent, $Y = S + M^*$, where the M^* was common to all stocks at a given time, the S independent for each stock Each of them obeyed the same diffusion law and corresponded to decisions on the value of stock, and the value of money We have thus satisfactorily accounted for the slopes of the lines in Figs 9 and 11, but we have not accounted for their absolute values, say the intercept for $\tau =$ one day

This intercept can be derived, at least in order of magnitude, from the discrete intervals of price and time in which equities are traded, mentioned under paragraphs 1 and 2 at the beginning of this paper

Let us approximate our condition $\mathcal{E}\Delta\log_e P = 0$ by supposing that there are just two possible changes of $\Delta\log_e P$ at each trade, up or down an amount h, h itself is unspecified at the moment It might be the logarithmic unit corresponding to ⅛ point, i e, $1/(8 \times \text{price})$, or it might be the spread/price At the end of k transactions, the semi-interquartile range for $\Delta\log_e P$ is, by the binomial distribution

$$\text{s q r}\ \Delta\log_e P = 0\ 6745\sqrt{k}\,h \qquad (17)$$

On a typical trading day, there are say 2×10^6 shares traded, or 2×10^4 round lots For 1000 issues traded this implies ≈ 20 round lots per issue, or between 10 to 20 transactions per individual issue, if we can suppose that the majority of transactions are for one- and two-round lots

Applying these figures to a typical, or median, stock price of \$40, if h corresponds to $1/(8 \times \text{price})$,

$$\text{s q r} = 0\ 6745\sqrt{15 \pm 5}/8\ 40 = 9\ 3 \times 10^{-3} \text{ or } 6\ 6 \times 10^{-3}$$

However, if h corresponds to the spread, estimated to be ½ point for a typical \$40 stock, then

$$\text{s q r} = 0\ 6745\sqrt{15 \pm 5}/2\ 40 = 7\ 7 \times 10^{-2} \text{ or } 3\ 7 \times 10^{-2}$$

The range corresponding to these two possibilities is plotted in Figs 9 and 10 Evidently a ⅛-point minimum change represents the dispersion developed at the end of one day better than our estimated typical spread However, we do not claim more than order of magnitude accuracy for the above calculation The spread would have given a better fit had we supposed more round lots per trade For the Canadian mines and oils, a diffusion rate based on spread might well be closer to the observation, and generally speaking individual decisions have to be based on a spread, not necessarily ⅛ point

THE STEADY-STATE DISTRIBUTION FUNCTION

Let us summarize the results of the above analysis of the data analytically $Y(\tau)=\log_e[P(t+\tau)/P_0(t)]$ can be expressed as the sum of two independent random variables $Y(\tau)=M(\tau)+S(\tau)$ $P_0(t)$ is the price of a random stock at a random time, t, $P(t+\tau)$ its price τ later M, the market or money variable is common to all stocks, S varies independently from one stock to the next In the above, and what follows, we no longer distinguish between M^* and its estimate M from a sample

The probability distribution of M is

$$\varphi(M)\,dM=dM\exp(-M^2/2\sigma_M^2\tau)/\sqrt{2\pi\sigma_M^2\tau} \tag{18}$$

Likewise, the distribution of S is

$$\varphi(S)\,dS=dS\exp(-S^2/2\sigma_S^2\tau)/\sqrt{2\pi\sigma_S^2\tau} \tag{19}$$

If τ is measured in years, $\sigma_M^2(\tau)=\sigma_M^2\tau$, and the variance developed at the end of unit time, or one year, is $\sigma_M^2=0\,0485$, from Fig 11, and similarly for the variance of S at the end of unit time, $\sigma_S^2=0\,0485$ from Figs 9 and 10 Thus we have taken $\sigma_S=\sigma_M$, though the value for σ_M is more uncertain, because of approximations in the reduction of the data This numerical value corresponds to a s q r of 0 148 at $\tau=1$ year, from fitting a line of slope ½ to all the data

The joint probability distribution of M and S is

$$\psi(M,\,S)\,dM\,dS=\varphi(M)\,\varphi(S)\,dM\,dS \tag{20}$$

The probability distribution of Y alone is

$$\begin{aligned}\varphi(Y)\,dY&=dY\int_{M=-\infty}^{\infty}\psi(M,S=Y-M)\,dM\\&=dY\exp(-Y^2/2\sigma^2\tau)/\sqrt{2\pi\sigma^2\tau},\end{aligned} \tag{21}$$

where $\sigma^2=\sigma_S^2+\sigma_M^2$ Note that in this form the variance of Y, $\sigma^2\tau$ is the sum of the variances of S and M, since Y refers to a random stock S, at a *random* time interval (for M) $\sigma_S^2\tau$ was the variance of Y from a sample for a single time interval, as determined in Figs 9 and 10 This distinction was noted following equation 16 The joint distribution of Y and M is

$$\begin{aligned}f(Y,\,M)\,dY\,dM&=\varphi(M)\,\varphi(S=Y-M)\,dY\,dM\\&=dY\,dM\exp[-½(1-\rho_{YM}^2)\{M^2/\sigma_M^2\tau\\&\quad-2\,\rho_{YM}\,YM/\sigma\sigma_M\tau+Y^2/\sigma^2\tau\}]/2\pi\sigma\sigma_M\,\tau\sqrt{1-\rho_{YM}^2}\end{aligned} \tag{22}$$

Here

$$\rho_{YM}^2=\sigma_M^2/(\sigma_M^2+\sigma_S^2),\qquad \sigma^2=\sigma_M^2+\sigma_S^2 \tag{23}$$

This is in a form that should be familiar to most biologists, and expresses mathematically facts, obvious to the casual student of the stock market There is a variable M called 'the market' which swings (is correlated with) the prices (Y) of all stocks up and down quite independently of circumstances peculiar to a given stock The correlation coefficient between the market changes and the price of a stock selected at random is $\frac{1}{2}^{1/2} = 0\ 707$, for $\sigma_M = \sigma_S$ This corresponds very roughly, in trader's parlance, to a 'sensitivity index' of unity [4]

LONG-TERM PRICE BEHAVIOR

Equation (21) has some rather interesting implications for the long term behavior of stock market prices, and when considered in conjunction with equation (20) it can be used to estimate the risks and degree of success of some elementary random investment strategies

Equation (21) gives the probability distribution of $Y = \log_e(P/P_0)$ where $P_0 = P_0(t)$, i e , of a randomly selected stock at a random time t, and $P = P(t+\tau)$, the price of the same stock τ later

The probability distribution of P itself is, from equation (21)

$$\begin{aligned} F(P)\ dP &= \varphi[Y = \log_e(P/P_0)](dY/dP)\ dP \\ &= \exp\{-[\log_e(P/P_0)]^2/2\sigma^2\tau\}\ dP/P\sqrt{2\pi\sigma^2\tau} \end{aligned} \tag{24}$$

The expectation value of P itself is

$$\mathcal{E}(P) = \int_0^\infty P\ F(P)\ dP = P_0 \exp(\tfrac{1}{2}\sigma^2\tau) \cong P_0(1 + \tfrac{1}{2}\sigma^2\tau) \tag{25}$$

The variance of P is

$$\sigma_P^2 = \mathcal{E}(P^2) - [\mathcal{E}(P)]^2 = P_0^2\ (\exp 2\sigma^2\tau - \exp\sigma^2\tau) \tag{26}$$

For τ small the dispersion is $\sigma_P \approx P_0\ \sigma\ \sqrt{\tau}$ Thus we see that the expectation value of log P does not change with τ, but the expectation value of P itself does At time $t+\tau$, the expected value has increased by $\frac{1}{2} P_0(t)\sigma^2\tau$, or not quite 5 per cent a year, using the numerical values for σ from equations (19)–(21) This discovery was first published by E C Smith and others have confirmed it [7]

This increase appears to be a quite satisfactory return for a random investment, or buyer's strategy—i e , buy at random and throw it in the box The difficulty with this strategy is shown in equation (26), which shows that the dispersion of P increases at a faster rate than $\mathcal{E}(P)$ Hence, one can never establish even a modest confidence interval around the mean to obtain, say an 80 per cent chance in any one transaction of realizing a profit The median of this distribution is still P_0 so that one-half of such investment choices result in a loss, which is more than compensated (in dollars of price, not $\log_e P$) by those which give a gain

The circumstance that the dispersion increases faster than the mean is not altered by using other sampling schemes, i e , by either diversification in stock (which would cut down the contribution of σ_S to σ_P) or by spacing the purchases in time, or diversification in the market variable, M, such as the monthly investment plan To achieve complete diversification in M, or time, one must use nonoverlapping intervals, which is only achieved in part by such schemes as dollar averaging

One should not suppose that the secular increase in time of stock market prices, indicated by equation (25), has anything to do with long-term inflation, or the growth of assets in a capitalistic economy This may be economic heresy, but the evidence, coupled with the argument below, seems to support this point of view We suspect rather that the evidence for inflation or deflation, if they exist, might be found by examining the skewness, or third moment of the distribution of $\Delta \log_e P$, but this is only a conjecture

The above discussion of equations (21) and (25) referred essentially to a buyer's strategy, the buyer being an individual who has money, and who makes decisions involving risk, with the expectation of acquiring more money Equations (23) and (24) offer identical opportunities to the seller, or individual who has stock and makes decisions involving risk with the expectation of acquiring more shares

In equation (24) $F(P)\,dP$ was the probability distribution of the number of dollars per share, to be expected from a sale in the future (at $t+\tau$) Hence, let $W=1/P$ be the number of shares acquired per dollar to be expected from a purchase in the future The probability distribution of W is

$$\begin{aligned} H(W)\,dW &= F(P=1/W)\,|dP/dW|\,dW \\ &= \exp[-(\log_e W/W_0)^2/2\sigma^2\tau]\,dW/W\sqrt{2\pi\sigma^2\tau}, \qquad (27) \\ W_0 &= 1/P_0, \end{aligned}$$

which is identical in form to the distribution function for P Hence everything we have said concerning the expected profit *in dollars*, the secular advance in prices, and their increasing fluctuation for the buyer who sells in the future holds with equal force for seller, who by holding his capital as cash can expect to acquire more shares by a purchase in the future Equations (25) and (26) hold for W as well as P unchanged

A NUMERICAL EXAMPLE

A simple numerical example may illustrate the symmetry of the opportunities presented, and the essential points Let us suppose shares of corporation X are currently selling (time, t) at \$100 per share It is estimated by both buyer and seller that at time $t+\tau$ in the future there is a fifty-fifty chance that X will sell at either ½ or twice (in general r or $1/r$) its present price, i e, the expected change in $\log_e P$ is zero The buyer, who has \$10,000 and would like more, can decide between holding his money, or buying 100 X's for future sale at time $t+\tau$, in which case the expected amount of his capital is $[(½)50+(½)200]100=\$12,500$ The expected amount of his capital for holding it is \$10,000, his present capital times one, a certainty

The seller who has 100 shares and would like more, can either hold it or sell and buy later In the latter case the expected number of shares is

$$(½)(10{,}000/50)+(½)10{,}000/200=125$$

If, contrary to the hypothesis of the Weber-Fechner law, buyer and seller *measure* their gains by number of shares and numbers of dollars, then the logical decision of the buyer would be to buy, and the seller to sell, and both would legitimately *expect* to come out ahead in re-acquiring what they originally possessed Both have to take a genuine risk Both could not gain as a result of the same single trial, but both could expect to do so for repeated, independent (hence not overlapping-in-time) trials

Under the hypothesis of the Weber-Fechner law, gains are *measured* by changes in logarithm of price, or logarithm of the number of shares In this case, the expected gain of each is zero Under these conditions there is what might be called 'indifference in the first order of decision' or logarithms of price changes are in a 'steady state of indifference' or statistical equilibrium between buyer and seller Whether or not they actually trade under these circumstances may well be determined by the secondary condition previously neglected in connection with the Weber-Fechner law These secondary conditions correspond, in the language of statistical mechanics, to the intermolecular forces that determine the nature of collisions, and of small departures from the steady state

This example illustrates almost all of the fundamental conclusions of this paper It shows the essence of risk-taking consequent to the expectation of a gain, how the gain should be measured, and the symmetrical properties of the stock market as a market both for stocks and money, as a fair meeting ground between buyers and sellers

REFERENCES

1 R J Trumpler and H F Weaver, *Statistical Astronomy*, University of California Press, Berkeley, 1953

2 S Chapman and T G Cowling, *The Mathematical Theory of Non-Uniform Gases*, Cambridge University Press, 1939

3 R C TOLMAN, *The Principles of Statistical Mechanics*, Clarendon Press, Oxford, 1938
4 R D EDWARDS AND J MAGEE, JR , *Technical Analysis of Stock Trends*, Stock Trend Service, Springfield, Massachusetts, 3rd Ed , 1954
5 A M MOOD, *Introduction to the Theory of Statistics*, McGraw-Hill, New York, 1950
6 A COWLES AND H E JONES, *Econometrica* **5**, 280 (1937)
7 C H DICE AND W J EITMAN, *The Stock Market*, McGraw-Hill, New York, 1952