(N)

PROFESSOR WESLEY MITCHELL ON INDEX-NUMBERS

[In this paper, published in the Economic Journal, 1918, under the title "The Doctrine of Index-Numbers according to Professor Wesley Mitchell," the plurality of conceptions attached to the term change-in-the-value-of-money, the variety of purposes subserved by an index-number for prices, is urged once more and with more confidence than before. That an index-number may be more than a register of a change in the value of certain specified articles, that there is an average trend of prices which may be expressed by methods other than those of a commercial account—this view is more acceptable now than it was thirty-five years ago. It is recognised in theory by Professor Mitchell, and realised in practice by Mr. Flux.]

The problem of which the object is to measure changes in the value of money has long exercised economists and statisticians. Thirty years have elapsed since the British Association appointed a committee for the purpose of investigating the best methods of ascertaining and measuring variations in the value of the monetary standard. The wording of this instruction may serve to remind us of the tremendous magnitude which the phenomenon to be measured has since the outbreak of war assumed. No one would now set out to ascertain the fact of a change in the value of money—a fact which in the peaceful eighties of last century could be disputed by sturdy mono-metallists without obvious absurdity. But though the fact now stares us in the face, the measurement of its magnitude is still important; perhaps more important than ever. For it can hardly be doubted that as the war goes on, and during the period of so-called "reconstruction," there will be required careful measurements of change in the purchasing power of money, with a view to the adjustment of wages and of other payments. And not only for practical purposes, but also in the interest of monetary theory, will such measurement be urgently required. In the controversies which will probably flourish in the early part of the twentieth, as in that of the nineteenth, century concerning the management of the currency during a great war, reference will certainly often be made to the index-numbers which represent the change from time to time in the level of general prices. If, as may be expected, the quantity theory of money is appealed to, it will be proper to construct another kind of index-number showing changes in the volume of trade. And other index-numbers there are which may be required in the course of reconstruction; in particular, those which measure wages nominal and real.

Coincidently with the increased demand for the use of index-numbers it is opportune that there has appeared a singularly comprehensive and lucid treatise on this species of measurement.[1] It is true that Professor Wesley Mitchell's monograph on index-numbers of wholesale prices does not cover all the ground which we have here in view. But the methods appropriate to the general problem can mostly be learnt from his discussion of a particular but leading case. That discussion is so complete and thorough that it almost dispenses the student who is not a specialist from the trouble of consulting the earlier literature of the subject. Within a limited but considerable and representative province Professor Mitchell has explored every inch of the ground. He has traced the many-branching paths which perplexed most of his predecessors. He has added clear directions showing where each of the paths leads.

The last-mentioned task is more difficult and important than may be supposed. It is a peculiarity of the problem that much thought must be expended in order to find the meaning of the question before you begin to answer the question. The practical man intent upon making or spending money does not suspect the ambiguity lurking under inquiries about its value. He asks what is the equivalent in our currency of the guinea in Charles II.'s time, and expects an answer as pat as if he had asked what is now the bank rate, or what the price of wheat. It is true that where the distance between the epochs compared is not so enormous, in the more usual comparisons of price-levels, the definition of the question is not so important; much the same answer may be given to different varieties of the question. The relation is like that between ethical theory and good conduct; if Bishop Butler and other moralists are right in thinking that

[1] Index-numbers of wholesale prices in the United States and foreign countries (*Bulletin of the United States Bureau of Labour Statistics*, 1915; whole number 173).

much the same conduct may follow from first principles so opposite as rational benevolence and rational self-love. When this analogy was suggested to Sidgwick, on the occasion of a meeting of the above-mentioned British Association Committee, the author of the *Methods of Ethics* made reply to the effect that, while frequently different methods might be adopted without obvious difference in practice, yet occasionally at critical turning-points the difference between opposite first principles would make itself felt decisively. We surmise that in like manner monetary distinctions which are otiose in ordinary times may have become significant under present conditions. All the greater is the debt of the economist to Professor Mitchell for having made the distinctions clear. Consider, for instance, Professor Irving Fisher's index-number in which each article is weighted in proportion to the number of times it is sold; quite properly, as Professor Mitchell points out (78),[1] with reference to Professor Fisher's purpose. In ordinary times there would probably be little difference between this number and that which is obtained by using the same commodities in the same quantities without taking account of the number of turnovers (Mitchell, *loc. cit.*).[2] But in war time, methods of business being considerably altered, it is possible that the distinction corresponds to a real difference.[3] The same may be said about another variety which Professor Mitchell thus distinguishes. "If the aim be merely to find the differences of price fluctuation characteristic of dissimilar groups of commodities, or to study the influence of gold productions, or the issue of irredeemable paper money upon the way in which

[1] The numerals in brackets refer to pages in Prof. Mitchell's treatise.

[2] Cp. *Memorandum British Association Report*, 1889. Professor Foxwell's method, above, p. 261.

[3] The subtlety of these distinctions deceives even experts. Thus the reviewer in the *Economist* (for January 12, 1918) criticising a recent publication in which it was held not to be proved, upon the lines of Irving Fisher, that money rather than goods was responsible for the rise of prices, observes triumphantly: "There can be no question that the increase in currency has been very much more rapid than the increase in the production of goods, unless we are to assume that this country, with four or five millions of its best men withdrawn into the Army, has been able to increase its production by more than 50 per cent." But the four or five millions have *not* been withdrawn from the production with which we are concerned in *this* inquiry. They (with their dependents) make, *primâ facie*, at least as great a pull as before upon the currency. Again the increase of women's and old men's paid work swells the denominator, which Irving Fisher calls "T." But then, asks the reviewer, why have prices risen? Quite conceivably, we reply, not so much because the quantity of money has increased out of proportion to the quantity of "goods" (in the sense here relevant), as because the circulation of the goods is less rapid (as suggested in the work criticised [*cp.* Lehfeldt, ECONOMIC JOURNAL, Vol. XXVIII. (1918), p. 111]). That is not "inflation" in the sense of causation on the side of money.

prices change, it may be appropriate to give identical weights to all the commodities " (78).[1] Again, the consumption standard, as based on family budgets, or more generally on the expenditure of the citizens in the way of consumption for the sake of personal or sympathetic satisfaction, exclusive of their collective expenditure on munitions for the satisfaction of patriotic motives, may well differ in war time from an index-number like that of Professor Irving Fisher, if there is included in the work which the currency has to do the payments by the Government for munitions. Conceivably, however differently from present experience, the *momentum* (price × velocity) of currency in relation to the " volume "—or, rather, the momentum, or *flow*—of goods, including munitions, might remain constant; while the prices of all the goods consumed by the citizen, exclusive of munitions, rose considerably.

Our readers are perhaps beginning to feel that they have had enough of this concept-splitting. Yet there remain certain varieties of index-numbers which we cannot pass over: two mentioned by Professor Mitchell and two which it did not come within his subject, more narrowly defined than ours, to mention. There is first the index-number intended to serve as a business " barometer " (66). If the aim be to construct a business barometer, the data should be prices from the most representative wholesale markets, the list should be confined to commodities whose prices are most sensitive to changes in business prospects and least liable to change from other causes, and the weights may logically be adjusted to the relative importance of the commodities as objects of investment. Professor Mitchell also directs attention to what he calls a " general-purpose " index-number, not adapted to any special end and in practice applied to very various purposes, of which more than a dozen are enumerated (26).[2] Professor Mitchell is no doubt right in thinking that " the day has not come when the uses of index-numbers are sufficiently differentiated and standardised to secure the regular publication of numerous special-purpose series." Till then " the users of index-numbers must put up with figures imperfectly adapted to their ends " (26).

[1] Cp. *British Association Memorandum*, 1887, section viii.: " Determination of an Index-number irrespective of the quantities of the commodities."

[2] Cp. *Memorandum*, 1887 (above, p. 255), " mixed modes, compounding the ends or means or several distinct methods " . . . " the most comprehensive . . purporting to be a compromise between all the modes and purposes—*the* method if practical exigencies impose the condition that we must employ one method, not many methods."

Another conception of the end, another definition of the value of money, is derived from Ricardo's axiom that "a commodity which at all times requires the same sacrifice of toil and labour to produce it is invariable in value." Professor Marshall has countenanced this view of our problem. In his evidence before the Precious Metals Royal Commission of 1888, speaking of the appreciation of gold,[1] he said: "When it is used as denoting a rise in the real value of gold, I then regard it as measured by the [increase] [2] in the power which gold has of purchasing labour of all kinds—that is, not only of manual labour, but the labour of business men and all others engaged in industry of any kind." It has been said that changes under this head are sufficiently reckoned with when the changes in average incomes are noted. This, however, may be questioned in time of a war involving enormous changes in the quantity of labour employed in production, additions here and subtractions there.

Nor can we pass over in silence Professor Nicholson's index-number based on capital.[3] It is remarkable that the conception which lies at the root of this method should have been that which, under a different aspect, first presented itself to Professor Lehfeldt in his independent and original investigation of the "absolute price of gold." [4] Professor Lehfeldt's *second* definition, referring to a "redistribution of effort of production" on the supposition of "the total of effort being unchanged," savours of the labour standard which we mentioned just now.

When we have decided what is the end at which to aim, we may go on to consider how the data are to be shaped to that end, and what data are to be sought. The step which is last in the analysis, as Aristotle would say, is first in the order of practice. The initial operation of collecting the original quotations of price requires more care and labour than might be supposed. "To judge from the literature about index-numbers, one would think that the difficult and important problems concern weighting and averaging. But those who are practically concerned with the whole process of making an index-number from start to finish rate this office work lightly in comparison with the

[1] Appendix to Final Report [C 5512] Question 9025. Quoted in the *British Association Memorandum* of 1889, p. 161.

[2] "Diminution" has been substituted for "increase" in the original by an obvious misprint.

[3] Described in the *British Association Memorandum*, 1887, section vi., above, p. 230.

[4] ECONOMIC JOURNAL (March, 1918, p. 108).

field work of getting the original data" (27).[1] The fathers of the English Statistical Society were so apprehensive lest the field work of collecting facts expressed in figures should be neglected if attention were diverted to drawing inferences from those facts that they proposed to divide the two kinds of work, and as the motto which they chose purported—*aliis exterendum*, surmounted by a wheatsheaf—themselves to gather in the harvest of statistics, while leaving it to others to thrash out the inferences. But Professor Wesley Mitchell has shown that it is possible for one and the same individual—combining official diligence with economic subtleties and statistical refinements—both to collect the raw material of primary data, and also to employ the complicated machinery which is required in order to render that material available for human use. We have not space to describe the excellent directions which are given to "the field worker collecting data for an index-number" (27). Indeed the whole of Part II., nearly two-thirds of the volume, dealing with index-numbers of wholesale prices in the United States and foreign countries, abounds with suggestions which may be useful to the practical statistician. Attention should be called to the suggestion that the facts may prove to be of more permanent interest than the theories which are now built thereon. "It is probable that long after the best index-numbers which we can make to-day have been superseded, the data from which they were compiled will be among the sources from which men will be extracting knowledge which we do not know enough to find" (30). We surmise that some of this future knowledge will be of the kind to which Professor Mitchell points: "to find how prices are interconnected, how and why they change, and what consequences each change entails" (29, 67).

Between the collection of the data and the completion of the index-number there are several intermediate processes which Professor Mitchell describes under the headings *base periods, the numbers and kinds of commodities included, problems of weighting, averages and aggregates*. We adopt this division, but we are not careful to follow the author's order as to the topics which are ranged under these four heads.

[1] One who was associated with Giffen when he was preparing the scheme of an index-number adopted by the British Association Committee can remember how much he was influenced in the selection of the items by the possibility of obtaining an available figure. He has himself expressed this in the second Report of the Committee (1888) which he drew up. "In dealing with the question practically those concerned must always have an eye upon the data, and consider what is practically attainable" (*loc. cit.* p. 183, and context).

Under the first head Professor Mitchell's most important contribution is the support which he gives to the method proposed by Professor Marshall, according to which the *base* adopted each year is constituted by the prices of the preceding year. "Chain" index-numbers it is proposed to call this species (36, 37, referring to 23). The ordinary "fixed base" index-number—for example, one constructed for the year 1913 with the prices of 1890–99 as base—is liable to an imperfection which is thus worded: "As the years pass by the commodities that have a consistent trend gradually climb far above or subside far below their earlier levels, while the other commodities are scattered between these extremes. Thus the percentages of variation for any given year gradually get strung out in a long, thin, and irregular line without any marked degree of concentration about any single point" (23). On the other hand, a careful scrutiny of the relative prices with which the "chain" method deals brings out the interesting circumstance that these percentages are grouped approximately according to the "normal law" of distribution. The familiar form which has been likened to (the front view of) a *gendarme's* hat reappears. But it should be noticed that in the centre of the hat there is a spike like that of a Prussian helmet—a "mode" which is very abnormal.

There is something impressive in the introduction of the normal law—the dominant principle of the higher statistics—into questions relating to money and prices. It is like the appearance of a distinguished savant as witness in a case relating to ordinary business. Let us make certain that the testimony is rightly interpreted.

When it is claimed as a merit of the "chain" data that they conform to the normal law, the question arises what advantage is there in such conformity. The feeling of statisticians on this question may perhaps be expressed by the old answer, "Si non rogas, intelligo." To reply that the law is convenient for purposes of calculation seems hardly relevant to the present inquiry. A deeper reason may be found in the presumption that the law is the outcome of numerous independent causes.[1] Since it is unlikely that independent phenomena should vary concurrently, we have here some guarantee of a certain stability in the grouping under consideration. It may be worth suggesting that prices regulated by Government in war time are determined by general

[1] Independence being understood in the sense explained by the present writer. *Journal of the Royal Statistical Society*, 1916, p. 462, and references there given.

rules rather than the plurality of fleeting causes which constitute the condition of the normal distribution. But we are not prepared to affirm that arbitrary governmental regulations will be deficient in the element of haphazard.

However, we do not dispute that it is a merit in a statistical group to conform to the normal law. We admit, too, that a continual elongation in one direction, such as Professor Mitchell has observed in the case of some relative prices, tends to deformation of the law. If *all* the prices behaved in this way, some moving upwards continually, the others downwards, they would

"leave in the midst a horrid vale,"

quite inconsistent with the normal contour. As nothing like this occurs, the continued elongations can be only moderately disfiguring; to what extent must be a matter of observation. We are not satisfied that this observation is performed with sufficient care by Professor Mitchell. He shows two diagrams, one representing the distribution of prices in 1913 as compared by the "chain" method with those of 1912, the other the distribution of 1913 compared with 1890–99 by the "fixed-base" method; and points out that the former set obey approximately the normal law. "But," he continues, "the distribution of the second set of variations (percentages of change from the average prices of 1890–99) . . . belongs to a different type. It has no pronounced central tendency; it shows no high degree of concentration around the arithmetic mean or median. It is more like an oblong than like the bell-shaped normal curve . . . its probable variation is five times as great as that of the corresponding variations for 1912 prices." This evidence is not conclusive; for it may be shown that the same appearance would be presented in like circumstances by the most perfectly normal distribution.

To construct an ideal distribution imagine a game in which each player moves a peg one step of, say, a quarter of an inch on a horizontal board either forward or backward, according as a tossed coin shows head or tail. A number of players that move thus east or west start from a line running north and south. Suppose that each player takes several steps in five minutes (corresponding to several changes in the price of an article during a year). At the end of that period the distances from the initial line will be ranged in at least rough correspondence to the normal law. Now let the race be prolonged for more than two hours—twenty-five periods each of five minutes. At the end of this time the normal distribution will be much more perfect (since more independent causes will have operated). But appearances would be against the new group. It might be said of it, as of the "fixed-base" series, that "it has no pronounced central tendency; it shows no high degree of concentration around the arithmetic mean," and so forth. For let there be given the "probable deviation" of the grouping after the first five minutes. Say it is so many quarters of an inch, or, better, a percentage of a certain standard length; namely, the number of inches which measures the distance of the initial line—the "career" of our imagined race—from a zero-point (west of that line). Let the said probable deviation be 4 (per cent.). Then the probable deviation for the group at the end of the race will be *five* times as great,[1] namely, 20. The latter grouping naturally appears "oblong," as contrasted with the "bell-shaped" contour of the former. The contrast is exhibited in the accompanying diagram,

16	38	60	60	38	16
14	15	15	15	15	14

where the notches on the line are placed at intervals of 4 (per cent. of the standard length), the probable deviation of the grouping formed by a five minutes' race. The upper figures show roughly (part of) a group of 240 observations thus formed. The lower figures show the grouping that may be with most probability expected after the forces tending to dispersion have acted for twenty-five times five minutes. It will be noticed that the shape of the bell is no longer conspicuous, about the centre at least. Certainly, if we exhibited the whole group, it might come out; it would come out if we represented, not the result of one race, but the average result of indefinitely numerous trials. But at any one race of the kind which we have described the grouping as a whole would assuredly appear rather "oblong." And yet it may be more normal than the small group. If the distance of each player from the starting-point at the end of the long race is divided by 5, the group so formed may be expected to comply with the ideal shape better than the set of points reached in the short race.

This contrast is not materially affected by the introduction of certain concrete circumstances. The coins which are used might be slightly unsymmetrical. There would then result a grouping which has many properties in common with the normal shape, the sub-normal curve as it has been called.[2] The sub-

[1] The *square root* of the number of independent constituents, which we have supposed to be twenty-five.

[2] See *Journal of the Royal Statistical Society*, "Mathematical Representation of Statistics," section iii., 1917, p. 65 *et passim*.

normal shape would persist, in spite of several modifications assimilating the case to that of price-variations. Then the causes tending to variation need not be perfectly independent. The steps need not be equal. There need not be several changes in each of the short periods. A variation may persist in one direction for several periods, provided that there is a chance of its being reversed. If, indeed, the last-named condition is removed on a large scale, the sub-normal character must disappear. Whether this or other abnormalities occur on such a scale as to vitiate the result is not to be decided off-hand, but by a careful scrutiny of the given statistics.

We have performed this scrutiny with respect to a set of 145 commodities selected by Professor Mitchell from the 241 above mentioned. We have examined the percentages presented by comparing the prices of 1913 with those of 1912 according to the "chain" method, and the percentages for 1913 according to the "fixed base" method with base 1890–99; and we are satisfied that the latter conform to the normal law at least as well as the former.[1]

But while thus questioning one of Professor Mitchell's premisses, we do not dispute his conclusion: "The longer a fixed-base system is maintained the more scattered become the relative prices as a rule" (37). Our discussion, however, warns us to accept with caution the corollary attached to this conclusion. "With a given body of quotations to build upon, chain relatives are more trustworthy than their rivals" (*loc. cit.*). Chain relatives relating to the preceding year are no doubt more trustworthy than their rivals when related to a much earlier period. But so are fixed-base index-numbers more trustworthy than a chain system if the former relate to the preceding year, the latter to a much earlier period.

The latter relation is dismissed too hastily in our judgment by Professor Mitchell. "Chain relatives for successive years . . . multiplied together to form a continuous series" (38) surely bring the later years into a relation with the earlier, which is as valid as most of the conceptions involved in an index-number?[2] We have performed this operation with respect to the chain index given by Professor Mitchell extending from 1890 to 1913; and have compared the result for 1913 with that which is given by

[1] The probable error for the group formed by comparison with a distant period, viz. nearly 20 (years), is to the probable error for the group formed by comparison between two consecutive years approximately in the ratio supposed in our illustration.

[2] Cp. *British Association Memorandum*, 1887, above, p. 219

the fixed-base method with base 1890–99. The difference appears at first sight marked; the median and the arithmetic mean according to the chain system being (for 1913) respectively 111·5 and 113, while, according to the fixed-base system, they are about 126 and 130 respectively. But it must be remembered that the chain starts from a greater height than the base of the fixed system; the level of 1890 (the base of the chain) being, with respect to the base of the fixed system, 113 or 114. This being taken into account, the consilience between the two systems is remarkable; and, indeed, greater than was to be expected.

It is tenable, we submit, that for certain purposes the chain system gives just as good a measure of the change in the price-level as the fixed-base system. For this reason we agree with Professor Mitchell that "it is an excellent plan to make from the original quotations two series of index-numbers—one a chain index and the other a fixed-base series."

Under the head of "Numbers and Kinds of Commodities Included" Professor Mitchell adduces a discovery for which students of his *Gold Prices under the Green-back Standard* [reviewed in the ECONOMIC JOURNAL, Vol. XVIII. (1908), p. 581] will be prepared. He has observed that the fluctuation in price from year to year is much greater for some kinds of commodities than for others (52 *et seq.*). Thus manufactured goods are steadier than raw materials. There are characteristic differences among the price fluctuations of the groups consisting of mineral products, forest products, animal products, and farm crops. Again, consumers' goods are steadier in price than producers' goods, the demand for the former being less influenced by vicissitudes in business conditions. Knowledge of this kind may be used to explain the discrepancies between different index-numbers which mix these classes of commodities in different proportions. Professor Mitchell bases on this observation a recommendation that the commodities utilised in the construction of index-numbers should be classified, not (or not only), as now, empirically, or with reference to practical interests, but (also) "upon causal lines, upon differences among the factors which determine prices, upon a principle of division which throws more light upon the workings of the complex system of prices."

In considering that system Professor Mitchell has thrown light upon the complex systems. For in the course of his observation he brings into view the interdependence or *correlation* between the prices of different commodities. There is a similarity between the price fluctuation of finished products and raw materials. This, however, is less than the similarity between the price fluctuations of finished products made from different materials.[1] The latter similarities, we surmise, are due to common causes, such as business cycles or changes in wages. The alternations of prosperity and depression no doubt affect all, or at least very many prices; but some much more than others. Thus the prices of minerals fluctuate with the alternations of business cycles better than the prices of other raw materials—farm and forest or animal products. Throughout the system there are found to be subtle correlations between observations which *primâ facie* are apt to be regarded as independent.

There is here exposed a feature which no doubt would be presented by other groups of statistics could they be as carefully examined. Observations seemingly independent are in reality honey-combed with correlations. Accordingly, calculations of probabilities based on the assumption of independence are apt to be inaccurate. Mathematical statisticians are too fond of calculating the " probable error " of averages on this assumption. They evolve, often with much labour and skill, a formula involving n, the number of observations, usually in the form of the *square root* of n as a factor of the denominator. They forget that commonly the given number n exaggerates the independence of the observations; owing to the existence of correlations, such as in the case of prices Professor Mitchell has so well expressed.[2]

Under the head " Problems of Weighting," Professor Mitchell propounds three questions: " Should the weights be sums of money or physical quantities? Should the weights be changed from year to year, or should they be kept constant? Should the weights be adjusted to the importance of the commodities as such, or should there be taken into account also the importance of the commodities as representing certain types of price fluctuations? " (78).

As to the first question, physical quantities measured by some

[1] The fact is thus happily expressed by our author : " As babies from different families are more like one another than they are like their respective parents, so here the relative prices of cotton textiles, woollen textiles, steel tools, bread, and shoes differ far less among themselves than they differ severally from the relative prices of raw cotton, raw wool, pig-iron, wheat, and hides."

[2] The great Laplace was not free from this assumption when he proposed to calculate the population of a country from the ratio between the number of baptisms and the population in different districts; and estimated the probable error of the calculation without taking into account the difference which no doubt prevails in respect of vital statistics between the inhabitants of different districts [cp. *Journal of the Royal Statistical Society*, Vol. LXXX. (1917), p. 549]. Probably only random samples such as those on which Dr. Bowley operates are quite —or at least very nearly—free from the influence in question (see Bowley, Presidential Address to Section F of the British Association, 1906; and *Livelihood and Poverty*, 1915).

conventional standard as a ton or a gallon are evidently improper weights for relative prices, ratios of which the type is p_r, the price of an article in the rth year, divided by p_0 the price of the same article in the base year (or period). But the *value* of the article (at the base, or some other suitable time) may properly be taken as the weight.[1]

As to the change of weights in view of change in the relative importance of commodities, Professor Mitchell points out that the compiler must choose between two evils, inaccurate weights and ambiguous price measures (79, referring to 31). With reference to fixed-base index-numbers, he considers that " the least objectionable compromise is probably to revise the scheme of weights, say, once a decade, and to show the effect of this change by computing overlapping results for a few years with both the old and new weights." [2] He puts " chain index-numbers " in a different category, for a reason that we have above questioned,[3] " since such series do not profess to yield accurate comparisons except between successive years " (80).[4]

Nor are we disposed to accept without qualification his answer to the third question, which is based on the following axiom. " An accurate measure of change in the level of all wholesale prices is not obtained unless all of the different types of fluctuation [referred to above, p. 394] . . . are represented in accordance with the relative importance of the commodities belonging to each." Very deep questions of first principle are here involved. We submit that the concept of an index-number for prices lies somewhere between two extreme definitions. One is the money value of a perfectly definite set of articles; for instance, a provision for certain functionaries of so much bread, sugar, uniforms, etc., from year to year (or at the same time in different countries). The sum total thus presented hardly deserves to be called an *index-number* (a title which, we observe with satisfaction, Professor Mitchell does not bestow as lavishly as some writers have done). Contrasted with a compilation which is of the nature of a commercial account is a true index-number, as described by Dr. Bowley: a quantity which we cannot observe directly, but which influences others which we can so observe.[5] The *quæsitum* thus conceived is related to the given price variations much as a physical quantity under measurement is related to a set of obser-

[1] *Cp.* below, par. 3.

[2] In illustration of this practice Professor Mitchell refers to Knibbs' *Cost of Living in Australia*, Commonwealth. . . Bureau . . . of Statistics Report, No. I., pp. xxiv, xlix.

[3] Above, p. 393.

[4] See *Elements of Statistics*, s.v. *Index number*.

[5] *Elements of Statistics*, chap. ix.

vations each purporting to represent the sought quantity. But the theory of errors-of-observation shows that in the combination of the given observations "less weight should be attached to observations belonging to a class which are subject to a wider deviation from the mean. Such would be prices of articles which, exclusive of the common price movement of all the selected articles, are liable to peculiarly large *proper* fluctuations." [1]

Ordinarily the required index-number is intermediate between the two extreme types which we have indicated. For even where the form is *primâ facie* a weighted sum—an aggregate of products each formed by multiplying a price by a quantity—still in our ignorance of the true factors the compound may assume the character of a true index-number.[2] Accordingly, a distribution of weight different from that which Professor Mitchell appears to prescribe would be ideally advisable. But where both the end to which our problem is directed and the means conducing thereto are so obscure and uncertain, we may acquiesce in our author's comment: "Perhaps it is a counsel of perfection to urge such requirements in systems of weighting."

It remains to consider the questions raised under the head "Averages and Aggregates." To some extent the answer to the questions under this and the preceding head will have been anticipated by the earlier discussion. For some decisions as to the scope and purpose of the index-number involve the choice of the method. Thus, if the purpose is that of Professor Irving Fisher, it follows at once that the proper combination of the data is the sum of the *values* (quantity $\times$ price) of the different commodities, each value weighted by the "turnover" of the commodity. The form of the index-number which purports to measure change in the cost of living is likewise predetermined. Ideally, at least, the form of a weighted sum is prescribed, though in practice it might be necessary to substitute some other.[3] The choice of average is wider in the case of other objects, of which some have been above mentioned.[4]

Professor Mitchell compares very fairly and fully the several available averages. Three stand out as selected candidates: the geometric mean, the median, and the arithmetic mean. With

[1] *British Association Memorandum*, 1887, p. 36. Cp. *Third Memorandum*, 1889, p. 157: "If more weight attaches to a change of price in one article rather than another it is not on account of the importance of that article to the consumer or the shopkeeper, but on account of its importance to the calculators of probabilities as affording an observation which is peculiarly likely to be correct."

[2] On this and other points connected with this discussion it may be allowable to refer to the present writer's *Lecture on Currency and Finance in Time o' War* Clarendon Press), 1917

[3] Above, par. 2.

[4] Above, p. 386 *et seq.*

regard to the geometric mean, Professor Mitchell points out—in addition to other considerations in favour of this form—that it is dictated upon a certain conception of the purpose in view. "If that purpose be to measure the *average ratio of change* in prices, the geometric mean is the best; indeed, in strictness, it is the only proper average to employ." [1] But, continues our author, "as a rule our interest does centre in the money cost of goods rather than in the average ratio of changes in price."

If the geometric mean is ruled out, it remains to weigh the rival claims of the median and the arithmetic mean. Professor Mitchell strikes the balance more impartially than the majority of practical statisticians. Still, we think that even he has not done full justice to the median. Its defects in respect of convenience and accuracy appear slightly exaggerated in his presentation.

An objection of the first kind is thus stated: "Medians of different groups cannot be combined, averaged, or otherwise manipulated with ease as can arithmetic means." For instance, the Bureau of Labour Statistics, after obtaining the sums of relative prices for farm products, clothing, etc., can obtain by simply summing up these sums the grand average for all commodities. But "it could not handle medians in this convenient fashion; instead of combining the sums from the groups, it would have to combine the single commodities." This objection is true and serious. But it is not in practice quite so serious as it seems in statement. In order to obtain the median of a composite group, one compounded of two groups for each of which the median has been found, it is not in general necessary to "combine the single commodities" in the sense of re-examining them all. It suffices to re-examine and rearrange those which are in the neighbourhood of the respective centres.

For example, here are two groups each consisting of *twenty-seven* observations ranged in the order of magnitude, which observations have each been obtained by adding together ten digits taken at random from mathematical tables.[2]

A.	27	30	31	32	33	34	36	37			40	41	*42*				46	47	48	49	50	51	52				59		64
		30										41	42				46						52						
		30																											
B.	29		31	32	33	34	36		38	39		41		43	*44*	45	46			49	50		52	53	56	57	59	62	
									38							45	46												
									38							45													
									38																				

[1] Jevons probably meant something like this by his somewhat obscure dicta as to the grounds for preferring the geometric mean. *Investigations*, pp. 23, 121.

[2] These figures are adduced with some comments relative to the present subject in the *British Association Memorandum* for 1889, p. 59. In the third group there given one of the sums, above 50, has been omitted by a misprint.

The median, being the *fourteenth* observation in the order of magnitude, is, for Group A, 42, and for Group B, 44. To find the median for the group compounded of these two we need not re-examine all the observations. For it is evident that the median of the compound cannot be greater than the larger of the two medians, viz., 44; nor less than the smaller, viz., 42. Accordingly, we may thus summarily, for the purpose in hand, re-write the data.

A.	XIII	42		XII	
		42			
B.	XII		43		XIV

Here the Roman numerals denote the number of observations which occur respectively above or below those given in Arabic figures. As the median of the composite group comprising *fifty-four* observations is intermediate between the twenty-seventh and twenty-eighth observations, it is evident at a glance that the required median is intermediate between 43 and 42, say 42·5. The process is easily extended to three or more groups.

However, we do not deny that the arithmetic mean has a considerable advantage over the median in virtue of the proposition, true of the former but not of the latter, that the mean of two (or more) means is the mean of the group formed by the constituents of both (or all) the several means.

Our difference with Professor Mitchell on another ground is more serious. He finds a difficulty in the use of the median in two opposite cases: when the given observations are either closely crowded, or widely dispersed about the centre of the group. As to the first case, it is said that the median may not answer precisely to its definition when several of the items to be averaged have identical values. For example, in Table II. [tabulating deviations presented by "chain" index-numbers] it often happens that the median falls on a large group of precisely identical figures, so that it ceases to be true that half of the cases are above and half below the median. Upon this it may be sufficient to say for the present [1] that the case in which there is an abnormal agglomeration about the centre is *primâ facie* one particularly favourable for the use of the median; since its probable error is less the greater, *ceteris paribus*, the height of the frequency curve at the middle.[2]

The opposite case of observations widely dispersed in the

[1] *Cp.* below, p. 404.

[2] More exactly, inversely proportional to the square root of the ordinate at the point on the abscissa where the median occurs. Cp. *Encyclopædia Britannica*, Art. "Probability," sect. 138, 139.

neighbourhood of the centre would be open to objection if the cause of the phenomenon were a depression in the form of the grouping, if the shape of the frequency curve about the centre resemble a valley between two eminences.[1] In that case, for the reason just now given, the probable error incident to the median would be particularly large. But this is clearly not the case contemplated by Professor Mitchell. He attributes the objectionable dispersion merely to the paucity of the observations. We shall therefore do no injustice to his argument if we suppose the grouping to be of an ordinary kind, in particular the normal law.

Upon this assumption it is at once to be admitted that the median is less accurate than the arithmetic mean, in so far as its probable error is a little greater, namely, in the ratio of about $1\frac{1}{4}$ to 1. That is all that we admit on the score of inaccuracy against the median. But we are by no means certain that we have apprehended Professor Mitchell's objections. Without being quite sure that we have located our author's position, we shall aim at three tolerably distinct points. (*a*) When there is a considerable interval between the position of the observation which forms the median and each of its nearest neighbours, then throughout a wide tract the position of the mean depends upon the accidents affecting a single observation. (*b*) The position assigned by the median is not perfectly definite. (*c*) The median is less responsive than the arithmetic mean to changes in the items.

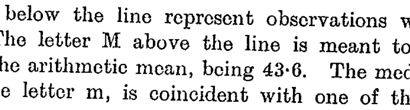
The first objection (*a*) is to be gathered from the following statements. "Where the numbers of items to be averaged is small, medians are erratic in their behaviour. . . . For in such groups there is often a considerable interval between the midmost relative price and the relative price standing next above it and next below. No change in any of the items, large or small, can alter the position of the median unless it shifts an item from the upper half of the list to the lower half or vice versa. But any change of this character, large or small, will make the median jump over the whole interval between its former position and that of the next highest or next lowest relative price, unless the change happens to place a new item within these limits" (85). Compare the dictum in the author's *Gold Prices . . . and Green-back Standard.* "The median . . . is rather erratic within limits of several points because its precise position is often dependent on the relative price of a single commodity which

[1] The supposition rejected as not appropriate to the data; above, p. 391.

stands in the middle of the scale of relative prices." [1] So again it is said: "When the numbers of commodities in the index-number is small, medians are likely to prove highly erratic, representing less the general trend of prices than the peculiarity of the data from which they are made" (90).

This objection is met by denying that the interval between two adjacent observations at the middle of the group is likely to be "considerable"; large relatively to the magnitude with which it is proper to compare that interval—that is, the *minimum mensurabile*, as we may say—that interval which is equal to (or of the same order as) the smallest degree which the compared method of measurement is capable of distinguishing with accuracy. For this minimum we may take at the least the "probable error" incident to the arithmetic mean. That the interval between adjacent observations is likely to be small compared with this minimum is sufficiently evidenced by the following proposition. When the number of observations (n) is large the interval at the middle of the group, which is as likely as not vacant, within which it is an even chance that no observation falls, is most probably very small compared with the probable range of the arithmetic mean (in the ratio of about $1 : \sqrt{n}$). When the number of observations is not large the proposition is less accurate. But it remains roughly true, as the number cannot be supposed very small consistent with the applicability of the theory of probabilities. Suppose, for instance, that the number of observations, is *twenty-five*, the number of a group which, according to Professor Mitchell, "illustrates the erratic character of the median." Then the space at the centre, which is as likely as not to be vacant, is about a quarter of the probable range to which the arithmetic mean is liable.

As a concrete example, let us take the groups above cited, formed by the addition of ten random digits. Here is reproduced the central region of Group A:—



The figures below the line represent observations which have occurred. The letter M above the line is meant to show the position of the arithmetic mean, being 43·6. The median, designated by the letter m, is coincident with one of the observa-

[1] *Op. cit.*, p. 58. Quoted with comment in the review of the work in the Economic Journal for 1898, Vol. XVIII., p. 581.

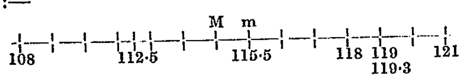
tions, 42. There is here, no doubt, a vacant tract between 43 and 46. But it is not considerable, regard being had to the roughness of the computation. For the probable error to which the arithmetic mean is liable is nearly 1·2; [1] that of the median is about 1·5.[2] Had the median occurred anywhere between 43 and 46 there would have been no reason for suspicion.

B.

Group B also does not countenance suspicion of the median. In its central tract here exhibited the largest gap is only 2; and the median, 44, as it happens, gives a better approximation than the arithmetic mean, 41·8, to the true value, which is 45.

If the object had been to ascertain whether there was any difference in the constitution of Groups A and B—whether B, for instance, had been constructed by the superposition of more or fewer digits than *ten* (the given number for A)—the median would give nearly as good an answer to this question as the arithmetic mean. The only difference is the one already acknowledged that the probable error of the median is a little larger. That difference would disappear if the number of observations (*in pari materia*) on which the median was based had been somewhat increased. The median of *forty* such observations would have afforded as good a test as the *arithmetic mean* of *twenty-seven*.

To give an example more germane to the present discussion, we have taken out, for the year 1890,[3] the relative prices that enter into the groups, numbering twenty-five each, which Professor Mitchell has instanced in connection with his remarks on the median. Here is exhibited the central portion of the first group or series. The arithmetic mean is at 115; the median at 115·5 :—

I. 

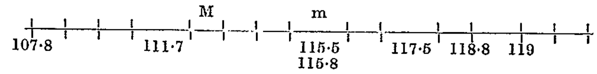
There is a vacant gap of 3 in the immediate neighbourhood of the median and of 4 not far from it. But what of that, seeing

[1] $= 0{\cdot}4769 \times \sqrt{\dfrac{10 \times 16{\cdot}5}{27}}$ (nearly).

[2] $= \sqrt{\dfrac{\pi}{2}} \times 1{\cdot}2$ (nearly).

[3] From Table II. of the *Bureau of Labour Statistics*, 1914; whole number 149.

that the probable error of the arithmetic mean is about 3, and that of the median itself greater !

II.


The central tract of the second series likewise does not bear witness against the median. The arithmetic mean is at 113; the median is at 115·5; the probable error about the same as for the first series.

(*b*) There seems to be expressed a fresh objection in the statement that it " is not always true of medians " that " their meaning is perfectly definite " (91). The meaning of the objection does not seem to us perfectly definite. Possibly it belongs to the class of difficulties apprehended in the case of numerous observations. Perhaps it is the same as the objection which we have already mentioned under that head. Perhaps it is the same as an objection which has been levelled by other statisticians against the median, viz. that it does not in general present a value so finely graduated as does the arithmetic mean. Consider, for example, the " race " above imagined, run by tossing coins—say a dozen every five minutes—and taking a step—say a quarter of an inch—forward or backward, according as each coin turns up head or tail.[1] If a number of players each proceed thus—starting from a starting-point labelled 100 (25 inches from zero)—at the end of the period the group will be distributed *discontinuously* at integral points. Now the arithmetic mean is not limited to integers, it may occur anywhere between two adjacent integers. But *primâ facie* the median *is* so limited, or, rather, it seems to be limited either to an integer if the number of constituents be odd, or to an integer + ½ if the number of constituents is even. The objection is not particularly applicable to the data with which our author is dealing, relative prices graduated to a tenth of 1 per cent. In any case, the objection is not very serious, since by a proper adjustment of the data a fractional value can be obtained for the median.[2]

(*c*) There remains to be considered the objection that " arithmetic means are more representative averages than medians, being affected by any change in the items of the group " (85) . . . " unlike medians, they [" aggregates "] allow every change in the

[1] Above, p. 391.

[2] See the present writer's articles *On the Use of Analytical Geometry to Represent . . . Statistics*, and *On the Mathematical Representation of Statistics. Journal of the Royal Statistical Society*, Vol. LXXVII. p. 732, LXXIX. p. 471, *et passim.*

price of every article to influence the result" (71). In making this objection, the author seems to have in view two groups *in pari materia* such that in passing from one to the other we find no change in the median, while there are changes among the other observations other than those determining the median, which changes affect the more sensitive arithmetic mean. Upon this it may be remarked that if there is this difference in the behaviour of the two averages, it is not to the credit of the arithmetic mean. The slight advantage which we have already allowed to the arithmetic mean would not be enhanced by this circumstance. Supposing that slight advantage corrected by basing the median on a greater number of observations, then the sensitiveness attributed to the arithmetic mean would be rather a defect than an advantage.[1]

But does the difference exist? Does the median, oftener than the arithmetic mean, does it even often, remain unchanged from one group to another? This may be doubted, if the data are finely graduated, or if graduation of the median by adjustment is practised.[2] The median seems, indeed, but only seems, to be irresponsive in certain circumstances—of perhaps frequent occurrence in the statistics of prices—which we shall indicate by continuing the parable of the indoor race.[3] Suppose that in the first five minutes several of the numerous players—late or dilatory—do not make a start, and that their positions at the end of the period are registered as being at the starting-point. Accordingly, at the end of a short period a good number of observations would be heaped up at the starting-point; the median would appear unmoved. But, of course, the position of those players who have not moved—whose position is not the result of steps determined by tossing coins—cannot be used to ascertain the asymmetry of the coins. For *that* purpose it would be proper to omit those dead-head observations, or to prolong the game until the slow players should come in. But for *other* purposes, of perhaps greater interest to the players, as relevant to the betting, it might be proper to take account of those nullities.

[1] The arithmetic mean in this respect might be compared with the method of examination by summing arithmetic marks practised at some public competitions as contrasted with examinations at one at least of our Universities where general unanalysable impressions have a due weight. The former method, no doubt, more frequently brings out candidates as unequal, but the distinction does not correspond to a real difference.

[2] Above, p. 403, note 2.

[3] Above, p. 391. Note that the spike-shaped "mode" there noticed is formed by prices which have *not moved at all* in the period under consideration; to be distinguished from those which have moved less than one mill.

Here, probably, is to be found the reason of the difference between Professor Mitchell and ourselves as to the worth of the median. We have been all along seeking to extricate from fallible observations a mean apt to represent the "general trend of prices" (9). That is the sort of index-number to which we submit that the median may be appropriate. But Professor Mitchell in his criticism of this average has presumably often in view some of the more directly practical purposes which have been distinguished, such as *par excellence* the determination of changes in the cost of living. For these purposes we at once admit that the median is not so appropriate as the combination of the kind which Professor Mitchell calls an "aggregate." [1] We entirely agree with him that "the best form for general purpose series is a weighted aggregate of actual prices."

[1] The term "aggregate" is felicitous as suggesting approach to that type which, as above explained (p. 396), is furthest removed from an index-number, the term least connected with the Calculus of Probabilities; infelicitous so far as it masks the affinity, not to say identity, between the proposed construction and the weighted arithmetic means used by Giffen, Palgrave, and the older statisticians (as to whom see *British Association Memoranda*, 1887, p. 264, and 1889, p. 139 *et seq.*). The words of Sidgwick there quoted: "Summing up the amounts of money paid for the things consumed at the old and the new prices respectively . . ." (*Political Economy*, Book I. ch. ii. section iii.), are appropriate to aggregates.