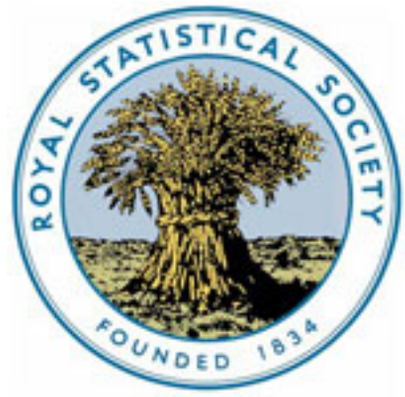

The Electrical Tabulating Machine.

By Dr. Herman Hollerith.

[Read before the Royal Statistical Society, 4th December, 1894.]

While engaged in work in the tenth census, that of 1880, my attention was called by Dr. Billings to the need of some mechanical device for facilitating the compilation of population and similar statistics. This led me to a consideration of the problems involved. I found, for example, that while we had collected the information regarding the conjugal condition of our 50,000,000 inhabitants, we were unable to compile this information even in its simplest form, so that, until the census of 1890, we never even knew the proportion of our population that was single, married, and widowed. Again, while we classed our population as native white, foreign white, and coloured, this was extremely unsatisfactory. For example, of what significance is it to know the number of children under 5 years of age who were native born? To have divided the native born into those of native parentage and those of foreign parentage, would have been practically impossible with the methods of 1880.

To obtain the population classified according to age, sex, and birthplace of mother could not have been considered. Again, it was apparent that if we wished to consider the progress of the negro in regard to illiteracy, we should know the number of illiterates at each age-period. In vital statistics much could be done in combining race, age, conjugal condition, occupation, and cause of death. Almost in every direction could be seen the need for combined or correlated statistics.

These were the considerations which prompted me to take up this problem, the result of which studies, after years of experimental work, are embodied in the apparatus or system which I will now briefly describe.

It must not be considered that this system is still in an experimental stage. Over 100,000,000 punched cards have been counted several times over on these machines, and this has afforded ample opportunity to test its capabilities.

I am glad to be able to say here to-day that in my struggle to secure the adoption of this system in the United States, I often had recourse, with great advantages, to references to and quotations from, the works of your Dr. Farr.

This system of electrical tabulation may perhaps most readily be described as the mechanical equivalent of the well-known

method of compiling statistics by means of individual cards, upon which the characteristics are indicated by writing. As it would be difficult to construct a machine to read such written cards, I prepare cards by punching holes in them, the relative positions of such holes describing the individual. In the United States Census we used cards of $3\frac{1}{4}$ inches by $6\frac{5}{8}$ inches, the surface of which was divided into 288 imaginary spaces $\frac{1}{4}$ inch square. To each of these spaces some particular value or meaning is assigned; a hole in one place meaning a white person, in another a black. Here a hole means a certain age-group, there it gives the exact year in that group. A combination of two holes in another part of the card indicates the occupation of the particular individual. In this way we not only recorded the answers to the twenty-six inquiries of the population schedule, but we also recorded the particular State, county, city, and enumeration-district in which the given person resided. Besides this, a number was stamped on the card, so that by these means any one of the 62,000,000 cards could be readily identified and compared with the original return.

This punching of the card, so far as the individual record is concerned, was done by means of the keyboard punch. The combination of holes representing the enumeration-district (over 40,000 in all), being the same for all the cards of a given district, was most readily punched by means of the gang punch. The punches being set for the given combinations, five or six cards were punched at one operation, the cost of this part of the work being thus relatively insignificant.

Having thus prepared a punched transcript for each individual, we are ready to tabulate them on the electrical machine. This consists primarily of a press or circuit-closing device, the upper and movable portion of which is provided with projecting spring-actuated needles, or points corresponding in number and relative position to the holes which may possibly be punched in the record card. The lower or fixed plate consists of a piece of hard rubber provided with a corresponding number of cups partially filled with mercury, which through suitable wires are connected with the binding posts of the switch board. If a punched card is placed in this bed, and the handle depressed, wherever there is a hole in the card the needle will dip down into the mercury, while at all other points the needles will be pressed back.

In connection with this so-called press counters are used. A counter consists of an electro-magnet, so arranged that each time a circuit is closed through it the armature is actuated so as to register 1. These counters can readily be re set to zero, and will count to 9,999.

If now we imagine such a counter connected to each mercury

cup, it is evident that if all the cards are successively placed in the press, the counters will ultimately give the total number of times any given hole occurred in the cards; or, in other words, a total showing the frequency of the different holes or items.

In practice, however, it is not sufficient to know simply the number of males and females, but we must know, for example, how many males there are at each age-period, as well as how many females at each age-period; or, in other words, we must count age and sex in combination. By a simple use of the well-known electrical relay we can secure this or any other possible combination. It must not be understood that only two items can be combined; in this way any number of items can be combined. We are only limited by the number of counters and relays.

As it would require 800 counters to compile a table of 800 columns, I have recourse to the use of a sorting box. This is simply a box divided into compartments (usually 24), placed by the side of the operator. The lids of these compartments are controlled by electro-magnets operated in exactly the same manner as the electro-magnets of the counters. If these magnets are connected to the mercury cups corresponding to age-groups, for example, and the cards are successively placed in the press, for each card a lid is automatically opened, according to the age of the individual represented by the given card. Each card having been deposited in the compartment opened by it, we have all our cards sorted according to twenty age-groups. If now each of these groups be passed through the machine provided with fifty counters, we obtain a result equivalent to a table of 800 columns.

It must also be noted that these two operations of sorting and counting can be conducted simultaneously, or either one independently of the other.

To show what can be done with such a machine, permit me to call your attention to the first handling of the punched cards of the United States Census. Here we obtained for each of the divisions of the population, *i.e.*, native with native parents, native with foreign parents, foreign white, and coloured, a classification according to sex, and the following age-periods: less than 1 year, 1 to 4, 5 to 9 10 to 17, 18 to 20, 21 to 44, 45 and over. At the same time we obtained a classification as to homes and farms, whether hired, owned free, or owned mortgaged. For the foreign, in addition to the above, whether a citizen or alien, and whether the person could speak English. For the coloured, a distinction as to Black, Mulatto, Quadroon, Octoroon, Chinese, Japanese, and Indian. This information it must be remembered was obtained for each enumeration district.

I will not weary you with the details of the various operations further than to call your attention to some few points.

We, I believe, pay about 35,000,000*l.* annually in pensions. How long this will continue is of course an interesting question. In case of any further legislation relating to pensions, we will know how many survivors of the late war there are at each age-period, also as regards the age of widows of soldiers. In other words, we now have some data upon which to base our calculations.

For certain classes of occupation, as many as ten distinct and different items were tabulated at one operation of the machine. Thus, for example, we now know the number in each occupation who were born in England, and who had mothers born in England.

It is of the greatest interest to know whether such a machine is accurate. Liability to error is of course always present, but with a properly arranged plan the possibility of an error going undetected is narrowed down to the one operation of punching. If the punched cards are verified, the subsequent operations can be fully checked by mechanical means. Even some of the errors of punching are detected by the electrical machine. If, for example, you forgot to punch whether the given foreign born person is an alien or citizen, the machine will not operate. Again, as this question is applicable only to males over 21 years of age, the machine takes all this into consideration, and does not refuse to count a female or a male under 21 even if citizenship is not punched. Only such errors of punching as are consistent will pass through the machine. I mean, for example, if a card is punched citizen when it should have been alien.

Naturally, in handling over 180 tons of cards as we did in the United States, there is apt to be some confusion. A few carpenters may get among the blacksmiths, or a few Bostonians may get mixed with the New Yorkers. These machines, however, were so connected that if while counting the blacksmiths of New York a carpenter should by accident come into the machine, it would not count. Likewise, if a Boston card came into the press it would be rejected.

Without the slightest delay such an electrical counting machine will read or test before tabulating whether the given person was white, native born, native father, native mother, male, blacksmith, and resident of new York City. If it agree in all these particulars, it would tabulate the person under from six to ten different items, whereas if the description did not tally, or any one of the required facts were not punched, the card would be rejected. An inspection of the card would then show the cause, and if due to an omission, this was supplied by reference to the original schedule.

Thus it is believed that the liability to error is much less, take it all in all, than with the old system of tallying, of ticking, or even the method of sorting and counting the original individual cards, as practised in Germany.

This system also possesses one advantage to which I would like to call your attention. In the case of current work the punched card may contain a pretty elaborate transcript which can be used for compiling the simpler monthly and annual tables, while at the end of say five or ten years there will be accumulated the cards, which will then be ready for an exhaustive tabulation. This would be true in case of registration of births, deaths, and marriages insurance experience, and many similar kinds of statistical work.

Discussion on MR. PORTER'S *and* DR. HOLLERITH'S *Papers.*

DR. W. OGLE congratulated Mr. Porter on having brought the gigantic task of the American census to a successful issue. He sympathised with him in his remarks on the reception often given to census results by the public. Those who found the figures tally with their expectations were given to cry out that a census was of little use, for they knew all about it beforehand; while those who found that the figures did not support their theories, attacked the census methods, and sometimes did not hesitate to accuse the census-takers of writing with an animus, and misrepresenting the facts.

He had no intention of comparing the results of the American and English enumerations; that would be an interesting but formidable undertaking. He would simply note that many of the phenomena presented by the population were the same in both countries. Such were the decreasing birth-rate, the decline in the rate of increase, and the migration of the rural population into the towns. These phenomena presented themselves both in America and in England, and indeed in other countries, and the explanation must therefore be sought not in conditions peculiar to ourselves, but in conditions shared by us with foreign States.

The American census differed enormously from the English census in its scope. Our own was limited to the enumeration of the people and houses, with some simple particulars concerning them, such as their age, sex, and occupation. But this was a comparatively insignificant part of the American census, which branched out into multitudinous inquiries, doubtlessly of high interest, but scarcely, as it seemed to him, coming properly under the designation of census work. For example, in the census report of 1881 there was, if he remembered rightly, a whole volume devoted to the habits and natural history of fishes, the

modes of capturing them, and other piscicultural matter. It appeared to him that it was a mistake to combine such inquiries as this with the census proper, that is, with the enumeration of the people. This latter could be carried out in a single day by ordinary enumerators at a comparatively cheap rate, while the special inquiries required a huge staff of highly-trained experts, a long period of time, and a fund which, it appeared from Mr. Porter's paper, ran into millions sterling. He was glad to see that Mr. Porter agreed with him that the two, as they required utterly different machinery, should be kept entirely separate; and he commended this opinion of Mr. Porter, as also his further opinion, that the questions in the enumeration schedule should be as few and as simple as possible, to those gentlemen who, in their laudable eagerness for information, tried at each successive census to introduce new topics of inquiry, and to add more and more questions to the householder's schedule. Mr. Porter had given an example of the hopelessness of getting trustworthy answers to any question that involved the least effort of thought or memory. What could be a simpler question, most outsiders would think, than this: *How many deaths have occurred in this family in the past year?* But Mr. Porter had found that some 50 per cent. of the deaths that had actually occurred were omitted in the answers to this simple question. In the English census of 1891 an addition, against which he had protested in vain, was made to the occupation questions in the schedule, and each person engaged in any industry was asked to state, by making a cross in the appropriate column provided for the purpose, whether he was employer, employed, or neither. In many schedules a cross was made in every one of the three columns, in many more in two of them, and, when the final tabulation was made, it was found that in some trades the persons returned as employers outnumbered those returned as employed. Another instance of the extremely untrustworthy manner in which ordinary men filled up a schedule was afforded by an inquiry some years back, on which he had been required to write a report, into the condition of working men in certain districts of London. The inquiry had been very badly conducted by the person who managed the business in one of the districts, and had to be made again a fortnight later. He had taken the trouble to compare the two schedules given by a number of men, and had found that the statements made by the same individuals as to their wages, and as to the time they had been out of work in the past six months, were in many cases utterly different.

He was as desirous as any one else to have ample statistics, but what he desired still more was to have accurate statistics; and the question, to his mind, of most practical importance at the taking of a census was, not how far they could persuade the authorities to extend the field of inquiry, but how they could manage to secure more trustworthy answers to the questions already included.

Mr. Noel A. Humphreys said that his experience of the last census almost made him long for another census in order to be able to try the effect of the electric tabulating machine. From the

description Dr. Hollerith had given, they must admit that it appeared to have already attained success, and in its present condition it promised greater accuracy in tabulation than the mechanical methods usually adopted, owing to the facility with which certain classes of errors could be detected by the machine. The value of census, and indeed of all statistical, work depended on the accuracy of the initial operations, whether by means of a tick made on an abstract sheet with the pen, or of a hole punched in a card. He was glad to find that there was nothing in the use of these cards to prevent the initial operations being rendered fully as accurate as the ticking system. The card punching had in fact the great advantage that the whole of the operation of the classification lies within a very moderate space immediately under the eye of the operator, instead of being spread over unwieldy abstract sheets such as have necessarily been used in the English Census Office. It must be remembered that with either method the value of the results depended upon the amount of trouble taken to verify this initial classification. One difficulty occurred to him in connection with the punching system, and that was, that all the varied classifications must be done by one person direct from the original schedule, whereas in the English Census Office the different processes are separated, in order to have experts in occupations to do one portion, experts in birth places to do another, and so on. In the American system the punching operation must be rendered entirely mechanical. This involved a system adopted to some extent in England, of "editing" or what might be called "coding" the returns before the cards are punched, in order to render this process purely mechanical. If the different portions of the returns, such as the occupations, were passed through the hands of an expert, and the difficult ones "coded" before being punched, the American process could be adopted without increasing the risk of inaccuracy. It is absolutely necessary, however, to insist upon the expenditure of the requisite time and money to re-working and verifying a sufficient proportion of the work. In the English Census Office an average of at least 10 per cent. was re-worked, in order to maintain a fair standard of accuracy, and to cancel really careless work. Dr. Hollerith had explained to him that this system was to some extent carried out at first in America, but that subsequently the operators (nearly all ladies) became so expert that the officials had felt themselves justified in reducing the amount to be re-worked. He himself believed that if in our next census Dr. Hollerith's machine could be tried, it would open out possibilities for an immense increase of valuable statistics, without any corresponding increase in the cost of the tabulation.

Mr. J. A. Baines, while joining in Dr. Ogle's congratulations, said that he could sympathise with Mr. Porter as to the want of a permanent census organization. In India he had none under him, but had been summoned by telegram to an empty office. He had in four or five months to draft his own rules, revising them chiefly in the light of the experience of some of his colleagues who had helped him ten years previously. His experience however

differed from that of Mr. Porter, in that the Indian Census was comparatively simple, except in point of numbers. The American Census resembled rather that of Rome, where it was not so much an enumeration of people as of property, for it was upon the results of these quinquennial enumerations that the taxation of the empire was based. In order to ensure a certain amount of accuracy, it was made a semi-religious function, and the Roman householder, after making the return, was required to repair to the temple and there solemnly take a bath, whence probably is derived the term "lustre"—a period of five years' intermediate washing being informal. As regards establishment, the great inconvenience of not having a permanent office was that several months had to be spent on each occasion in teaching a large number of assistants. Mr. Porter had endorsed Lord Farrer's opinion that a statistician must have "views," and he (Mr. Baines) concurred in this, so long as the man did not hold settled convictions: he must have an open mind, be loyal to the *data*, and not use the figures merely to support conclusions already formed. As regards the method of enumeration, the census in India was almost everywhere taken, as in England, in a single night, and the inquiries were consequently confined to much narrower limits than in the United States. The great advantage of this was the far more accurate enumeration, which necessitated simple questions. He did not think that errors in defect balanced those in excess; they might do so over very large areas, but they certainly did not in tracts of from 750,000 to a million persons, as he had carefully tested this point in India. They had a great deal to learn from America, and he had often wished to have the Hollerith machine for his census work, although it was doubtful whether he could have found anyone quick enough to learn to use it, India being very backward in taking up such inventions, apart from a difficulty about the current. Five years ago there was hardly a single typewriter even in any office in the country. But for the abstraction and tabulation he had been able to draw on a large staff, as India possessed a literary proletariat fairly educated for that sort of work. Their caste prohibited them from taking any manual employment, and he had therefore found the full staff he required at very short notice, and they worked, when trained, like machines. In the first few days 90 per cent. of their abstractions had to be worked over again but they were soon able to reduce this quantity, until they finally re-worked only one-third of the results, and even less. But when it came to the tabulation, he had to weed out the office very much, because he found that the most efficient abstractors were by no means good tabulators. They were proficients in the mechanical process, but could not combine the figures, and were constantly "miscolumning," as they termed it. Still, on the whole, he had derived an enormous advantage from this staff of high caste literary candidates. The cost of operations subsequent to enumeration amounted to about Rs. 6·5 per 1,000 of the population, out of a total of Rs. 11, and was thus cheaper than the census in almost any other country, even when allowance is made for the comparative cheapness of living in India. He was glad to find that

Mr. Porter's views as to the enumeration of occupations coincided with his own. It was perfectly impossible, in his experience, to get a correct census of occupations, either by means of enumerators who filled up the schedule, or by trusting the householders to do so. Mr. Porter had mentioned the difficulty regarding languages in some large towns, and he might remark that his schedules, weighing in all 290 tons, were printed in seventeen languages, and filled up with the names of over one hundred dialects. On one other point he was completely in accord with Mr. Porter, and that was in regard to simplicity of the schedule. If ever they attained to a quinquennial census, the aspiration of every census commissioner, the schedule should not be one bit more elaborate than the one suggested in the paper, and even in that two or three columns might possibly be omitted.

Mr. F. B. GARNETT said that he, as well as all others who had attempted it, had found great difficulty in endeavouring to estimate the total value of property in the United Kingdom, and had found that the only method of reaching even an approximate total was by capitalising the annual value of real property of various kinds, profits of trades or professions, and income from all other sources at so many years' purchase, while the number of years might vary at different times, according to circumstances; and those who entered on such speculations were not always agreed as to the proper estimate to be taken for a basis in respect of each class of property. He would like to ask, therefore, whether in the United States they possessed advantages which did not exist in England for arriving at what might be considered a true valuation of the national wealth.

Mr. JOHN B. MARTIN reminded the audience that the Royal Statistical Society had always been most persistent in its endeavours to secure a quinquennial census. It had appointed a special Census Committee, and only last summer a deputation from the Council had waited on the President of the Local Government Board, the Right Hon. G. J. Shaw-Lefevre, an ex-President of the Society, and discussed with him the possibility of bringing about that most desirable consummation. It was agreed on both sides that such an event would entail the institution in Great Britain of a skeleton permanent department such as Mr. Porter proposed for the United States. He reminded the audience also of the existence of the International Statistical Institute, founded under the auspices of this Society, and comprising the most distinguished statisticians of all countries, the aim of which was to bring the methods of the statistics of various countries into line. There was some shadowing forth of progress in this direction in the paper just read, and he hoped that the time would come quickly when uniform schedules of vital statistics and of census returns would be adopted throughout the whole civilised world.

The CHAIRMAN (Dr. G. B. LONGSTAFF), in rising to propose a very cordial vote of thanks to Mr. Porter and Dr. Hollerith, said

that the point which Mr. Porter had urged with so much force, as to the institution of a permanent census office, was one which that Society had constantly brought before the Government. As regarded a quinquennial census they had now achieved a substantial success. The scheme for the partial equalization of rates in London depended upon a system of levying an even rate over the whole metropolis and distributing it according to the population of the several districts, and the Government had at last realized that this could only be done by a frequent census. It was accordingly now ordained that a census should be taken in London in 1896, under the auspices of the London County Council, and paid for out of the rates. He trusted that Dr. Hollerith's machine could be utilised on this occasion, and that by this means the census could be made at the same time cheaper and more comprehensive. There was one point to which Mr. Porter had not alluded. He believed that the *fons et origo* of Mr. Porter's trouble was the constitutional difficulty in the central Government's collecting information. He hoped that Mr. Porter would enter on the arduous task of reforming the constitution of the United States at least to the extent of enabling them to have a permanent census office and permanent records. If the registration system could be established throughout the United States, it would prove a gigantic gain to statistical science. As a theatre of statistical investigation at the present time, no country in the world was so interesting as the United States, but with the imperfect tools at command they were unable to deal with a number of problems quite unequalled elsewhere for magnitude and interest.

Mr. PORTER replied to various questions. With regard to an inquiry by the Chairman, he had no reason to believe that there was much error in the amount of mortgages in consequence of people withholding statements. It had at first been feared that difficulties might arise in this connection, owing to the enumerator being a local man, and the latter had therefore made only the inquiries noted on the schedule; the remainder of the information was obtained by direct correspondence from Washington. Besides this precaution, anyone refusing to give information under this head was liable to a fine of 100 dollars, and in some cases the penalty was actually enforced. Still, there were doubtless errors in the returns. He had at the outset protested strongly against the inquiry, but his opinion had been over-ruled. Replying to Major Craigie and others concerning the definition of a mortgage, Mr. Porter said that the investigation included what were known in America as real-estate mortgages, and not chattel-mortgages. All mortgages in America, to be valid, were recorded, but it often happened that people, when they had paid off a mortgage, forgot to see that it was taken off the record; and many which had been partly paid off were found to be still upon the record. It was not possible therefore to base the results simply upon the records. For full explanations of the methods adopted to prosecute this inquiry, he referred to the paper by Mr. George K. Holmes published in the *Journal* of the Society (vol. lvi, September, 1893).

Replying to Mr. Garnett, he thought that in America they had rather better grounds for estimating the national wealth than in England. Taxation in every State and territory was based on the valuation by assessment of both personal and real estate. In arriving at the figures, they first ascertained the value of every given township in the country, and then addressed about 150,000 circulars to all real-estate owners and county recorders, requesting them to fill up certain schedules as to the value of property in the locality, and if possible to show the actual value at which any piece of property had changed hands. In this way they obtained a sort of guiding basis for each district. The value of the railroads, &c., was then included as indicated by Mr. Garnett, but he quite agreed with him as to the difficulty of assessing trade profits. At the best these statements were but estimates, and must be so regarded, even though great care had been taken in their preparation. The seventh volume of the Tenth Census contained a full statement of the plan adopted by him (Mr. Porter) at that time, and a similar, though perhaps more elaborate scheme was used in the Eleventh Census.

Sir Francis S. Powell inquired as to the method of taking the religious census. Mr. Porter said that he had avoided seeking information from the individual. Inquiry was made of the officials of the various religious bodies throughout the whole country. The 21,000,000 who were returned in this connection represented only the communicants belonging to each body, while no enumeration had been made from house to house, it was quite possible to form an estimate of the religious population of the United States upon the basis of the communicants reported sufficiently accurate for all purposes. The usual way of computing religious population was by multiplying the number of communicants of any Protestant denomination by $3\frac{1}{2}$. This is on the supposition that for every communicant there were $2\frac{1}{2}$ adherents, including young children. Dr. Carroll's figures show the Protestant population to be 49,630,000, and the Catholic population at 7,362,000. This, he contends, stands for the Christian population of the country. After allowing for Jews, he arrives at the conclusion that about 5,000,000 of our population in 1890 belong to the non-religious and anti-religious classes, including free-thinkers, secularists, and infidels. In collecting the statistics of religions, he had met with a very good instance of the necessity of simplicity in the inquiry. In reply to the question, "What is the seating capacity of your church?" the answer, in many cases was, "Oak," "Pine," &c., as the case might be.

Mr. HOLLERITH, with reference to the liability of the machine to get out of order, said that in making the contract with Mr. Porter for the use of the machines, a clause had been introduced involving a penalty of 10 dollars to be paid by the inventor for every day that the machine was out of order. The penalty had never been enforced. Replying to other questions, the cards were only passed through the machines four times during the whole of the operations. They were stored for future reference,

and every card was so marked that it could be compared with the original schedule at once. An error card had occasionally been met with, and had been sent to the "error" section of the department, where a few men were employed rectifying mistakes.

He should be happy to exhibit the machine in operation upon some future occasion. With regard to the system itself, he had been making many experiments upon a considerable scale, involving the use of about 5,000,000 cards, on which the amounts, instead of individual items, were recorded. One card, representing a farm, for instance, would give the total number of acres, number of horses, value of the farm, and the implements, stock, produce, &c. By attaching to the machine an integrating device, he had been able to make it a species of adding machine. This led directly into other fields; he was now experimenting on railroad accounts, and hoped to accomplish this also.

A cordial vote of thanks was passed to Mr. Porter and Dr. Hollerith, and the proceedings terminated.