"The average Contribution of each several Ancestor to the total Heritage of the Offspring." By FRANCIS GALTON, D.C.L., Sc.D., F.R.S. Received and Read June 3, 1897.

In the following memoir the truth will be verified in a particular instance, of a statistical law of heredity that appears to be universally applicable to bisexual descent. I stated it briefly and with hesitation in my book 'Natural Inheritance' (Macmillan, 1889; page 134), because it was then unsupported by sufficient evidence. Its existence was originally suggested by general considerations, and it might, as will be shown, have been inferred from them with considerable assurance. Consequently, as it is now found to hold good in a special case, there are strong grounds for believing it to be a general law of heredity.

I have had great difficulty in obtaining a sufficient amount of suitable evidence for the purpose of verification. A somewhat extensive series of experiments with moths were carried on, in order to supply it, but they unfortunately failed, partly owing to the diminishing fertility of successive broods and partly to the large disturbing effects of differences in food and environment on different broods and in different places and years. No statistical results of any consistence or value could be obtained from them. Latterly, while engaged in planning another extensive experiment with small, fast-breeding mammals, I became acquainted with the existence of a long series of records, preserved by Sir Everett Millais, of the colours during many successive generations of a large pedigree stock of Basset hounds, that he originated some twenty years ago, having purchased ninety-three of them on the Continent, for the purpose. These records afford the foundation upon which this memoir rests.

The law to be verified may seem at first sight too artificial to be true, but a closer examination shows that prejudice arising from the cursory impression is unfounded. This subject will be alluded to again, in the meantime the law shall be stated. It is that the two parents contribute between them on the average one-half, or (0·5) of the total heritage of the offspring; the four grandparents, one-quarter, or $(0{\cdot}5)^2$; the eight great-grandparents, one-eighth, or $(0{\cdot}5)^3$, and so on. Thus the sum of the ancestral contributions is expressed by the series $\{(0{\cdot}5)+(0{\cdot}5)^2+(0{\cdot}5)^3$, &c.$\}$, which, being equal to 1, accounts for the whole heritage.

The same statement may be put into a different form, in which a parent, grandparent, &c., is spoken of without reference to sex, by saying that each parent contributes on an average one-quarter, or $(0{\cdot}5)^2$, each grandparent one-sixteenth, or $(0{\cdot}5)^4$, and so on, and that generally the occupier of each ancestral place in the nth degree, whatever be the value of n, contributes $(0{\cdot}5)^{2n}$ of the heritage.

In interbred stock there are always fewer, and usually far fewer, different individuals among the ancestry than ancestral places for them to fill. A pedigree stock descended from a single couple, m generations back, will have 2^m ancestral places of the mth order, but only two individuals to fill them; therefore if $m = 10$ there are 1024 such places; if $m = 20$ there are more than a million. Whenever the same individual occupies many places he will be separately rated for each of them.

The neglect of individual prepotencies is justified in a law that avowedly relates to average results; they must of course be taken into account when applying the general law to individual cases. No difficulty arises in dealing with characters that are limited by sex, when their equivalents in the opposite sex are known, for instance in the statures of men and women.

The law may be applied *either* to total values or to deviations, as will be gathered from the following equation. Let M be the mean value from which all deviations are reckoned, and let D_1, D_2, &c., be the means of all the deviations, including their signs, of the ancestors in the 1st, 2nd, &c., degrees respectively; then

$$\tfrac{1}{2}(M+D_1)+\tfrac{1}{4}(M+D_2)+\&c. = M+(\tfrac{1}{2}D_1+\tfrac{1}{4}D_2+\&c.)$$

It should noted that nothing in this statistical law contradicts the generally accepted view that the chief, if not the sole, line of descent runs from germ to germ and not from person to person. The person may be accepted on the whole as a fair representative of the germ, and, being so, the statistical laws which apply to the persons would apply to the germs also, though with less precision in individual cases. Now this law is strictly consonant with the observed binary subdivisions of the germ cells, and the concomitant extrusion and loss of one-half of the several contributions from each of the two parents to the germ-cell of the offspring. The apparent artificiality of the law ceases on these grounds to afford cause for doubt; its close agreement with physiological phenomena ought to give a prejudice in *favour* of its truth rather than the contrary.

Again, a wide though limited range of observation assures us that the occupier of each ancestral place *may* contribute something of his own personal peculiarity, apart from all others, to the heritage of the offspring. Therefore there is such a thing as an average contribution appropriate to each ancestral place, which admits of statistical valuation, however minute it may be. It is also well known that the more remote stages of ancestry contribute considerably less than the nearer ones. Further, it is reasonable to believe that the contributions of parents to children are in the same proportion as those of the grandparents to the parents, of the great-grandparents to the grandparents, and so on; in short, that their total amount is to be expressed by the sum of the terms in an infinite geometric series diminishing to zero. Lastly, it is an essential condition that their total amount should be equal to 1, in order to account for the whole of the heritage. All these conditions are fulfilled by the series of $\frac{1}{2}+\frac{1}{2}^2+\frac{1}{2}^3+$ &c., and by no other. These and the foregoing considerations were referred to when saying that the law might be inferred with considerable assurance *à priori*; consequently, being found true in the particular case about to be stated, there is good reason to accept the law in a general sense.

The Bassets are dwarf blood-hounds, of two, and only two, recognised varieties of colour. Excluding, as I have done, a solitary exception of black and tan, they are either white, with large blotches ranging between red and yellow, or they may in addition be marked with more or less black. In the former case they are technically known and registered as "lemon and white," in the latter case as "tricolour." Tricolour is, in fact, the introduction of melanism, so I shall treat the colours simply as being "tricolour" or "non-tricolour;" more briefly, as T. or N. I am assured that transitional cases between T. and N. are very rare, and that experts would hardly ever disagree about the class to which any particular hound should be assigned. A stud-book is published from time to time containing the pedigrees, dates of birth, and the names of the breeders of these valuable animals. The one I have used bears the title 'The Basset Hound Club Rules and Stud-Book,' compiled by Everett Millais, 1874–1896. It contains the names of nearly 1000 hounds, to which Sir Everett Millais has very obligingly, at my request, appended their colours so far as they have been registered, which during later years has almost invariably been done. The upshot is that I have had the good fortune to discuss a total of 817 hounds of known colour, all descended from parents of known colour. In 567 out of these 817, the colours of all four grandparents were also known. These two sets are summarised in Table I and discussed in Table V, and they afford the data for Tables II, III, and IV. In 188 of the above cases the colours of all the eight great-grandparents were known as well; this third set is discussed in Table VI.

Partly owing to inequality in the numbers of the tricolours and non-tricolours, and partly owing to a selective mating in favour of the former, the different possible combinations of T. and N. ancestry are by no means equally common. The effect of this is conspicuous in Table I, where the entries are huddled together in some parts and absent in others. Still, though the data are not distributed as evenly as could be wished, they will serve our purpose if we are justified in grouping them without regard to sex; or, more generally, if we treat the 2^n components of each several A_n, whatever be the value of n, as equally efficient contributors.

Our first inquiry then must be, "Is or is not one sex so markedly prepotent over the other, in transmitting colour, that a disregard of sex would introduce statistical error?" In answering this, we should bear in mind a common experience, that statistical questions relating to sex are very difficult to deal with. Large and unknown disturbing causes appear commonly to exist, that make data which are seemingly homogeneous, very heterogeneous in reality. Some of these are undoubtedly present here, especially such as may be due to individual prepotencies combined with close interbreeding. For although this pedigree stock originated in as many as ninety-three different hounds, presumably more or less distant relations to one another, some of them proved of so much greater value than the rest that very close interbreeding has subsequently been resorted to in numerous instances. In order to show the danger of trusting blindly to averages of sex, even when the numbers are large, I have compared the results derived from different sets of data, namely from those contained in the last two columns of Table I, where they are distinguished by the letters A and B, and have treated them both separately and together in Table II. They will be seen to disagree widely, concurring only in showing that the dam is prepotent over the sire in transmitting colour. According to the A data, their

ERRATUM.

Vol. 61, p. 410.

In the memoir read June 3, 1897, entitled "The average Contribution of each several Ancestor to the total Heritage of the Offspring" the words *sire* and *dam* have been accidentally transposed in Table II, and consequently, in the deduction therefrom at the bottom of page 404, the latter should be that, in the present case, the sire is more potent in transmitting colour than the dam, in the ratio of 6 to 5. This error does not affect the general conclusions of the memoir, because the ratio of 6 to 5 was treated as an insignificant disproportion, and the two sexes were dealt with on equal terms.

FRANCIS GALTON.

relative efficacy in this respect is as 58 to 51, say 114 to 100; according to the B data, it is as 47 to 32, say 147 to 100. Taking all the data together, it is as 54 to 45, say 120 to 100, or as 6 to 5.

It does not seem to me that this ratio of efficacy of 6 to 5 is sufficient to overbear the statistical advantages of grouping the sexes as if they were equally efficient, the error in one case being more or less balanced by an opposite error in the other. It is true that the reciprocal forms of mating are by no means equally numerous, the prevailing tendency to use tricolours as sires being conspicuous. Still, as will be found later, on the application of a general test, the error feared is too insignificant to be observed. Should, however, a much larger collection of these data be obtained hereafter, minutiæ ought to be taken into account which may now be disregarded, and the neglect of female prepotency would cease to be justified.

The law to be verified supposes all the ancestors to be known, or to be known for so many generations back that the effects of the unknown residue are too small for consideration. The amount of the residual effect, beyond any given generation, is easily determined by the fact that in the series $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}$, &c., each term is equal to the sum of all its successors. Now in the two sets of cases to be dealt with the larger refers to only two generations, therefore as the effect of the second generation is $\frac{1}{4}$, that of the unknown residue is $\frac{1}{4}$ also. The smaller set refers to three generations, leaving an unknown residual effect of $\frac{1}{8}$. These large residues cannot be ignored, amounting, as they do, to 25 and 12·5 per cent. respectively. We have, therefore, to determine fixed and reasonable rules by which they should be apportioned.

The requisite data for doing this are given in Table III, which shows that 79 per cent. of the parents of tricolour hounds are tricolour also, and that 56 per cent. of the parents of non-tricolour hounds are tricolour. It is not to be supposed that the trustworthiness of these results reaches to 1 per cent., but they are the best available data, so I adopt them.

It will be convenient to use the following nomenclature in calculation:—

a_0 stands for a single member of the offspring.

a_1 for a single parent; a_2 for a single grandparent, and so on, the suffix denoting the number of the generation. A parallel nomenclature, using capital letters, is:—

A_0 stands for all the offspring of the same ancestry.

A_1 for the two parents; A_2 for all the four grandparents, and so on. Consequently A_n contains 2^n individuals, each of the form a_n, and A_n contributes $(0{\cdot}5)^n$ to the heritage of each a_0; while each a_n contributes $(0{\cdot}5)^{2n}$ to it.

In the upper part of Table IV the ratios are entered of the average contributions of T. supplied by *known* ancestors. Nothing further need be said about these, except that they are styled coefficients because they must be multiplied into the total number of offspring, in order to calculate the number of them that will, on their separate and independent accounts, be probably tricolours.

We have next to explain how the coefficients for the *unknown* ancestry have been calculated, namely, those which are entered in the lower part of Table IV. Suppose all the four grandparents, A_2, to be tricolour, then only 0·79 of A_3 will be tricolour also, $(0{\cdot}79)^2$ of A_4, and so on. These several orders of ancestry will respectively contribute an average of tricolour to each a_0 of the amounts of $(0{\cdot}5)^3 \times (0{\cdot}79)$, $(0{\cdot}5)^4 \times (0{\cdot}79)^2$, &c. Consequently the sum of their tricolour contributions is

$$(0{\cdot}5)^3 \times (0{\cdot}79) \{1 + (0{\cdot}5) \times (0{\cdot}79) + (0{\cdot}5)^2 \times (0{\cdot}79)^2 + \&c.\},$$

which equals 0·1632. The average tricolour contribution from the ancestry of *each* of the four tricolour grandparents must be reckoned as the quarter of this, namely, 0·0408.

By a similar process, the average tricolour contribution from the ancestry of *each* non-tricolour grandparent is found to be 0·0243.

When the furthest known generation is that of the great-grandparents, the formula differs from the foregoing only by substituting $(0{\cdot}5)^4 \times (0{\cdot}79)$ for $(0{\cdot}5)^3 \times (0{\cdot}79)$. This makes the average tricolour contribution from the ancestry of the whole eight tricolour great-grandparents equal to 0·08160, and that from the ancestry of *each* of them to be one-eighth of this, or 0·0102.

In a similar way the tricolour contribution from the ancestry of *each* non-tricolour great-grandparent is found to be 0·0061.

The following example shows how the coefficients in Table IV were utilised in calculating the general coefficients entered in Table V.

2 Parents, T_1 (personal)	0·5000
3 Grandparents, T_2 (personal)	0·1875
1 Grandparent, N_2 (personal)	—
3 Grandparents, T_2 (ancestral)......	0·1224
1 Grandparent, N_2 (ancestral)	0·0243
Total tricolour contribution ..	0·8342

The coefficient 0·83 will consequently be found under the appropriate head in Table V, where the total number of offspring ("all cases") is recorded as 119. By multiplying these together, viz., $0{\cdot}83 \times 119$, the "calculated" number of 99 is obtained. It will be seen that the observed number was 101, a difference of only 2 per cent.

The extraordinarily close coincidence throughout the two tables, V and VI, between calculation and observation, proves that the law is correct in the present instance, and that the principle by which the unknown ancestry was apportioned, is practically exact also. It is not so strictly exact as it might have been, because the whole of the available knowledge has not been utilised. The 0·79 applied to A_4, &c., requires some small correction according to the known colours of the *offspring* of A_3. If they had been all tricolour the 0·79 would have to be increased; if all non-tricolour, it would have to be diminished. Having insufficient data to check a theoretical emendation, I note its omission, but shall not discuss the matter further.

It will be easily understood from these remarks how *collateral* data are to be brought into calculation, for if the collaterals were more tricolour than the average of hounds, the 0·79 would have to be somewhat increased (but not beyond the limiting value of 1·00); if less tricolour than the average, the 0·79 would have to be diminished. The knowledge of collaterals would be superfluous, if that of the direct ancestry were complete, but this important prolongation of the present subject must not be considered further on this occasion.

There are three stages in Tables V and VI at which comparisons may be made between calculated results and observed facts.

(1) *The Grand Totals.*—In Table V the sum of all the calculated values amounts to 391; that of all the observed ones to 387, which are closely alike. In Table VI they amount to 180 and 181 respectively, which is a still closer resemblance. Consequently the calculations are practically exact *on the whole*, and the error occasioned by neglect of sex, &c., is insignificant.

(2) *The Subordinate Pairs of Totals.*—These are entered at the sides of the tables, and are nine in number, namely, 236, 239; 149, 139; 6, 9; 53, 56; 52, 56: 9, 9; 8, 6; 49, 46; 9, 8. The coincidences are striking, in comparison with such results as statisticians have usually to be contented with; the second pair, 149, 139, is the least good, and will be considered in the next paragraph.

(3) *Individual Pairs of Entries.*—There are 32 of these; here also calculation compares excellently well with observation, excepting in the line that furnishes the "subordinate totals" of 149, 139, where the "all cases" of 37, 158, 60 yield the tricolour contingents of 20, 79, 36. Dividing each tricolour by the corresponding "all cases," we obtain what may be called "Coefficients from Observation," to compare with the calculated coefficients. They are as follows:—

		Diff.		Diff.	
Coefficients from observation	54	(−4)	50	(+10)	60
" " calculation	66	(−8)	58	(−7)	51

The great irregularity of the entries in the upper line shows that the observed values cannot be accepted as true representatives of the normal condition. I have not unravelled the causes of this error, and it is not urgent to do so, since its ill effects are swamped by the large number of successes elsewhere.

In order to satisfy myself that the correspondence between calculated and observed values was a sharp test of the correctness of the coefficients, I made many experiments by altering them slightly, and recalculating. In every case there was a notable diminution in the accuracy of the results. The test that the theory has successfully undergone appeared on that account, to be even more searching and severe than I had anticipated.

It is hardly necessary to insist on the importance of possessing a correct law of heredity. Vast sums of money are spent in rearing pedigree stock of the most varied kinds, as horses, cattle, sheep, pigs, dogs, and other animals, besides flowers and fruits. The current views of breeders and horticulturists on heredity are contradictory in important respects, and therefore *must* be more or less erroneous. Certainly no popular view at all resembles that which is justified by the present memoir. A correct law of heredity would also be of service in discussing actuarial problems relating to hereditary longevity and disease, and it might throw light on many questions connected with the theory of evolution.

Table I.—Pedigrees of Parental and Grand Parental Colours (Tricolour (T) or Non-Tricolour (N)).

The Sex of the Ancestors is taken into Account, but not that of the Offspring.

Sire's sire / Sire's dam			T T	T T	T N	T N	T T	T T	T N	T N	N T	N T	N N	N N	N T	N T	N N	N N	Totals.	Others of which all G. P.'s are not known.
Dam's sire / Dam's dam			T T	T N	T T	T N	N T	N N	N T	N N	T T	T N	T T	T N	N T	N N	N T	N N		
Sire.	Dam.	Offspring.	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*	*n*	*o*	*p*	A.	B.
Tric.	Tric.	Tric.	106	38	47	12	12	7	2	4	3	..	3	4	..	..	..	..	239	87
		Non-T.	13	4	12	2	2	1	1	..	..	..	..	3	..	..	..	..	38	20
Tric.	Non-T.	Tric.	..	..	16	3	..	..	1	..	3	..	..	..	1	1	..	..	25	15
		Non-T.	..	..	6	1	..	..	3	..	2	..	..	4	1	1	..	..	18	17
Non-T.	Tric.	Tric.	20	45	4	20	11	4	4	..	..	2	..	3	1	..	..	..	114	30
		Non-T.	17	50	16	9	5	..	1	1	..	8	.	1	1	..	..	..	109	64
Non-T.	Non-T.	Tric.	..	..	..	2	..	..	..	..	..	5	..	1	..	1	..	..	9	11
		Non-T.	..	..	..	8	..	..	..	4	..	3	..	..	..	..	..	..	15	6
Totals			156	138	101	57	30	12	12	9	8	18	3	16	4	3	..	..	567	250

Table II.—Offspring of one parent Tricolour (T) and of one Non-tricolour (N). The sex of the parents is not regarded.

From Table I.		Observed.			Per cents.	
		Tricolour.	Non-tricolour.		Tricolour.	Non-tricolour.
Sire T, dam N	A	114	109	223	51	49
	B	30	64	94	32	68
	Sum	144	173	317	45	55
Dam T, sire N	A	25	18	43	58	42
	B	15	17	32	47	53
	Sum	40	35	75	54	46

Table III.—Distribution of T and N colour in Parents, when the Offspring are T and N respectively.

From Table I.		No. of T offspring.			Parents* of T offspring.	
Sires.	Dams.	A.	B.	Total.	T.	N.
T	T	239	87	326	652	0
T	N	25	15	40	40	40
N	T	114	30	144	144	144
N	N	9	11	20	0	40
Totals				530	836	224
Per cent. of parents*				50	79	21

From Table I.		No. of N offspring.			Parents* of N offspring.	
Sires.	Dams.	A.	B.	Total.	T.	N.
T	T	38	20	58	116	0
T	N	18	17	35	35	35
N	T	109	64	173	173	173
N	N	15	6	21	0	42
Totals				287	324	250
Per cent. of Parents*				50	56	44

* More properly "Parental Places"; the number of these, though not that of the individual parents, being always double the number of any group of offspring.

Table IV.—Tricolour coefficients.

	Ancestry known up to and inclusive of	
	Grandparents.	Great-grandparents.
Personal allowance of T for each		
Tricolour parent	0·2500	0·2500
„ grandparent	0·0625	0·0625
„ great-grandparent	—	0·0156
(No allowance for Non-tricolours.)	—	—
Ancestral allowance of T for each		
Tricolour grandparent	0·0408	—
Non-tricolour „	0·0243	—
Tricolour great-grandparent	—	0·0102
Non-tricolour „ „	—	0·0061

Table V.—Calculation and Observation Compared.

The pedigrees are utilised up to the second ascending generation. Sex not taken into account.

No. of tricolours in parents.		Number of tricolours in grand-parents.				Total tricolour offspring.	
		4	3	2	1	Calculated.	Observed.
		a	*bcei*	*dfgjkm*	*hlno*		
2	All cases	119	119	28	11		
	Coefficient	0·91	0·83	0·76	0·68		
	Tricolour calc'd.	*108*	*99*	*21*	*8*	*236*	
	„ observed	106	101	24	8		239
1	All cases	37	158	60	6		
	Coefficient	0·66	0·58	0·51	0·43		
	Tricolour calc'd.	*24*	*92*	*30*	*3*	*149*	
	„ observed	20	79	36	4		139
0	All cases	..	..	18	6		
	Coefficient	..	..	0·26	0·18		
	Tricolour calc'd.	..	..	*5*	*1*	*6*	
	„ observed	..	..	7	2		9
					Grand totals	*391*	387

Table VI.—Calculation and Observation Compared.

The pedigrees are utilised up to the third ascending generation. Sex not taken into account.

No. of tricolours in parents.	No. of tricolours in grand-parents.		Number of tricolours in great-grandparents. 8	7	6	5	4	Total tricolour offspring. Calculated.	Total tricolour offspring. Observed.
2	4	All cases	2	25	14	16			
		Coefficient	0·96	0·94	0·92	0·90			
		Tricolours calc'd.	*2*	*24*	*13*	*14*	..	*53*	
		,, observed	2	25	14	15	..		56
	3	All cases	..	18	21	16	6		
		Coefficient	..	0·87	0·85	0·83	0·81		
		Tricolours calc'd.	..	*16*	*18*	*13*	*5*	*52*	
		,, observed	..	17	19	14	6		56
	2	All cases	..	3	2	3	3		
		Coefficient	..	0·81	0·79	0·77	0·75		
		Tricolours calc'd.	..	*2*	*3*	*2*	*2*	*9*	
		,, observed	..	2	2	3	2		9
1	4	All cases	..	2	1	9			
		Coefficient	..	0·69	0·67	0·65			
		Tricolours calc'd.	..	*1*	*1*	*6*	..	*8*	
		,, observed	..	1	..	5	..		6
	3	All cases	1	28	14	31	9		
		Coefficient	0·64	0·62	0·60	0·58	0·56		
		Tricolours calc'd.	*1*	*17*	*8*	*18*	*5*	*49*	
		,, observed.	1	16	12	8	9		46
	2	All cases	..	..	4	13			
		Coefficient	..	..	0·54	0·52			
		Tricolours calc'd.	..	..	*2*	*7*	..	*9*	
		,, observed	..	..	1	7	..		8
							Grand totals ..	*180*	181

The summed data derived from Table IV, form the coefficients entered in Tables V and VI. These are multiplied into the corresponding number of "all cases," and the result gives the "calculated" number of tricolour hounds among them.

The entries of "all cases" and of "tricolours observed" in Table V are deduced from Table I, by combining the appropriate columns. The letters at the top show which columns are combined.

Seven other observed cases, disposed in three groups, are scattered beyond the limits of Table VI; two of these seven cases are tricolour.