DISCUSSION

of locating a paper on 'corrosion of aircraft' while looking for interconnexions between aeronautics and chemistry. Did he not assume at least an elementary subject categorization in recognizing that 'corrosion' is a part of 'chemistry'? Either it is necessary for the index to make this recognition and thus choose 'chemistry' as one of the descriptors, or else it is necessary for the retriever to realize that when he wants all papers on chemistry he must also look under 'corrosion'. It would seem that some sort of categorization, either explicit or implicit, cannot be avoided. It might be well to get away from the limitations of the categorical tree structure by allowing multiple lines to be drawn upwards as well as downwards in the hierarchal schemes (i.e. allowing one descriptor or subject to belong to more than one parent category).

C. N. MOOERS in reply: Exceedingly simple information retrieval machines can and do perform their function at high speeds and come out directly with abstracts, as for example the card sorting system of reference 1. It was the point of my paper to examine, through a study of syntax and coding, how far the capabilities of such machines (or simple ones like them) could be extended into retrieval of structured information. The Minicard[2] is not a simple machine, not all of which have yet been built and tested. Its semantics and syntax are still hypothetical. Its sorting speed of 1,800 items per minute is modest for a machine of its complexity, being only about twice that of the simple machine mentioned above. From reports, it appears that it will be an excellent specialized machine for the storage of items on film snippets and their high speed selective photographic copying. For collections of greater than one million items, the Minicard searching time of about nine hours is rather long, thus limiting its applicability to information retrieval.

The matter of 'subject categories' for descriptors is very important and it would take several full papers to treat[3]. The use of any hierarchal classification tree like the Universal Decimal Classification hinders, rather than aids, information retrieval. It restricts one's retrieval questions to only those entries that fit in the classification tree. The scheme for using 'multiple lines drawn upwards and downwards' seems to be close to Perry's 'abstraction ladders', and as such is a sort of super classification tree. The list of all these ladders must be set down before you can begin any coding by the scheme, and there is a fantastic number of possible ladders. There is the interesting question of what to do where the lines cross (as they do almost everywhere), and this has never been discussed. My preferred solution to Mr. Irving's question is to use a short written schedule of descriptors (a few hundred in any discipline), and to have the retriever use this schedule as a check list in formulating his search. In the schedule he will find both 'chemistry' and 'corrosion'. His decision of how to use the concepts would then come from the explanation written for each of the descriptors in a list of 'scope notes'. On rare occasions descriptors are actually related by an inclusion relation, but when they are, only two levels of hierarchy are allowed, and only a small fraction of the descriptors are so limited.

REFERENCES

1 MOOERS, CALVIN N. 'Zatacoding Applied to Mechanical Organization of Knowledge', American Documentation, 2 (1951) 20
2 TYLER, A. W., MEYERS, W. L. and KUIPERS, J. W. 'The Application of the Kodak Minicard System to Problems of Documentation', American Documentation, 6 (1955) 18
3 MOOERS, CALVIN N. 'Scientific Information Retrieval Systems for Machine Operation—Case Studies in Design', Zator Technical Bulletin No. 66, Boston, 1951

15

ON THE LANGUAGE OF TAXONOMY: AN OUTLINE OF A 'THERMOSTATISTICAL' THEORY OF SYSTEMS OF CATEGORIES WITH WILLIS (NATURAL) STRUCTURE*

BENOÎT MANDELBROT

University of Geneva

THE NUMBERS of species within genera in natural systems of categories appear frequently to be ruled by a probability distribution first observed by J. C. WILLIS[1]. This distribution can be considered to be an approximation to an *exceptional stable* distribution of Cauchy-Paul Lévy. Its study then runs quite parallel to that of the *normal stable* distribution of Laplace-Gauss that is found in thermodynamics. However, specific properties of the distribution are quite different and remarkable, and they can be a basis for a wealth of models for the rules of formation of higher taxonomic categories, on natural Linnaean classification systems.

THE PROBLEM

A Linnaean classification system, or taxonomic tree, is largely an arbitrary method of identifying species, by successive dichotomies, irrespective of any assumption about the existence or the values of species frequencies (however, for comparison purposes, frequencies will sometimes be assumed to exist, in the sequel, and to be equal). Besides, some, largely arbitrary, intermediate steps of the identification are given special names, as 'genera', 'families', 'orders', or still higher 'categories'.

It is found that different genera in a family usually contain a very variable number of species. Very small or very large genera are frequent, and there is very little clustering around some median value, with superposed fluctuations, which is 'normal' for physical quantities. This inequality is of very great interest to the information theorist who, in the absence of any other data, will be shown to be able to derive structural laws of the whole tree from laws ruling the partition of species among genera. These laws somewhat reduce the arbitrariness of taxonomy, and show that natural systems are in many ways 'random' or 'extremal'.

DATA

WILLIS[1], considering the accepted taxonomies of a few very large biological 'families', has observed that the number $g(s)$ of genera, having each exactly

* The mathematical developments included in the paper presented at the Symposium have been omitted from this version. It is hoped that they will soon appear in full detail in *Information and Control*, a quarterly journal.—Ed.

s species (s not too small), can consistently be represented by what we shall call 'Willis's relationship'

$$g(s) = Ps^{-(\alpha+1)}$$

The 'intensive' parameter α is always $0 < \alpha < 1$, and is usually close to 0·5. Examples are given in *Figure 2*, taken from YULE[2]. Willis's relationship holds, with various values of α, for families taxonomized by 'splitters' as well as by 'lumpers'; that is, by taxonomists favouring 'rather small' or 'rather large' genera. It seems therefore safe to conjecture (though it needs to be checked) that a lumper's and a splitter's taxonomies for the same family differ only by the value of α. Further, the lumper's genera would be called subfamilies by the splitter. It can be conjectured that the numbers of the splitter's genera within a lumper's genus also can be represented by Willis's relationship (or 'follow' Willis's law).

This relationship is not limited to biology; it was found by ZIPF[3] to hold for generic categories of business catalogue items, for generic names of professions etc. It seems to be very widespread, but is not believed to hold for any of the non-Linnaean systems of categories of mineralogy, meteorology etc.

Willis probability distribution

It is believed that, in order to understand Willis's relationship, one should consider that a random process is at the root of the choice of the taxonomic trees, of genera (and even of species), and so also of the number of species in a genus. We shall therefore conjecture that Willis's relationship is the frequency distribution of a sample drawn from a random population. The probability distribution of this population should then be close to the 'Willis distribution'

$$p(s) = Ps^{-(\alpha+1)} = K(\alpha)\alpha^s s^{-(\alpha+1)}(1 \le s < \infty; K^{-1}(\alpha) = \Sigma_1^\infty s^{-(\alpha+1)})$$

Goodness of fit was checked by YULE[2]. [Notice that, before estimating α one should apply 'A. M. Turing's correction'; that is, replace all s by $s^0 = (s+1)g(s+1)/g(s)$; see Good[4].] Since Willis's relationship does not hold for s small, $P^{-1} \ne K^{-1}(\alpha) = \Sigma_1^\infty s^{-(\alpha+1)}$ and P is an independent parameter. We write P as $K(\alpha)\alpha^s$ in order to introduce a new ('extensive') scale parameter a, proportional to the median value $\hat{s}$ of s (that is, to the value of s as likely as not to be exceeded, since this value is given by $K(\alpha)\alpha^{-1}\alpha^s \hat{s}^{-\alpha} = \frac{1}{2}$). However, if $a < 1$, the above formula introduces the spurious possibility of void genera ($p(O) > 0$) and if $a > 1$, the domain of variation of s must be restricted to $1 < k(a) \le s < \infty$.

The striking fact about $s(= \Sigma_1^\infty s^{-\alpha})$ is the slowness of its decrease with s^{-1}; the expected value of $s(= \Sigma_1^\infty s^{-\alpha})$ is infinite, as well as the variance, so there is no sense in speaking of 'fluctuations around some mean value'. These properties will induce quite unexpected behaviours. (These would disappear if Willis's distribution could be replaced by the distribution $p(s) = P^s \exp(-bs)s^{-(\alpha+1)}$, with b very small, which has some theoretical justification. However, a system with such a structure would have none of the deep homogeneity properties of a system following a Willis distribution.)

'Willis systems' of categories will either be families in which species within genera follow Willis's law, or hierarchies of categories in which items of each level follow the law relative to the items of the level above it.

A modified Willis probability distribution
$p'(s)$ is defined as the coefficient of u^s in the series development of the 'generating function' (a kind of 'spectral distribution'):

$$G(u) = \Sigma_0^\infty p'(s)u^s = 1 - a^\alpha(1-u)^\alpha$$

if s is large, $p'(s) \sim a^\alpha s^{-(\alpha+1)}/\Gamma(-\alpha)$, which behaves exactly like the $p(s)$ of Willis's probability distribution proper.

PROPERTIES OF WILLIS SYSTEMS OF CATEGORIES

It can be shown that, from the fact that the expected value of the number of species in a genus is infinite, it follows that Willis systems have quite exceptional properties relative to operations of addition, division and multiplication, as defined below. Besides, if the number of species in a family be S, the expected number of genera will be $G = \sin \alpha\pi S^\alpha / P\pi = R(\alpha)S^\alpha$ (FELLER[5]). It is seen that the number of genera per species varies like $S^{\alpha-1}$, and tends to zero as the size of the family tends to infinity, $S \to \infty$.

Addition: Lumping together of independent Willis categories (LÉVY[6,7])
The *sum* of very many, I, Willis variables of same α does not tend to (is not attracted by) the usual 'normal' stable Laplace-Gauss random variable. But this sum if divided by $I^{1/\alpha}$, instead of the usual $\sqrt{I}$, tends to an 'exceptional' stable Cauchy-Paul Lévy random variable, for which the distribution $F(x)$ is not known in closed form (except if $\alpha = 0.5$) but one knows the 'characteristic function' (another kind of 'spectral distribution')

$$\varphi(t) = \int_0^\infty \varepsilon^{ixt} dF(x) = \exp\left[-a\left\{1 + i\left(\frac{tg}{|t|}\frac{\alpha\pi}{2}\right)\right\}t/|t|\right]|t|^\alpha$$

Clearly, if x_1 and x_2 are independent and stable, so is their sum. The Willis distribution itself is already a good approximation to a stable distribution, especially for $\alpha \sim 0.5$. (However, it presents the opposite defect that, whereas one knows $p(s)$, one does not have any closed formula for its characteristic function $\varphi(t)$ (or its generating function $G(u) = \int_0^\infty u^s dF(x)$ $= \varphi(-i \log u)$). A further approximation, having closed forms for both distribution and spectrum, is provided by the Modified Willis distribution. From the approximate stability of the Willis distribution, it follows that, if the number of species in a category is the sum of I independent random numbers x_i following Willis (α) distributions with parameters a_i—except for the frequencies of small values of the variable—then the total number in the category follows a Willis (α) distribution with $a = (\Sigma a_i^\alpha)^{1/\alpha}$—except for the frequencies of small values of the number: this exceptional zone increases with the exceptional zones of the addends, and with their number. If the a_i are equal, $a = a_i I^{1/\alpha}$, therefore the median number of species, which is proportional to a, increases like $I^{1/\alpha}$. For example, if a family contains G genera, the median number of species in it varies like $G^{1/\alpha}$, so that the median number of species per genus tends to infinity with G, like $G^{1/\alpha-1}$.

Division: Splitting of Willis categories: Microcanonical components of a genus
Suppose that one wishes to *split* each genus of a Willis system, through some additional 'feature'. The presence or absence of such a feature may be ruled

by some deterministic process (*e.g.* equal splitting, or splitting into a fixed (*small*) number of species and the rest). It may also be ruled by a random process, independent or not from that which determined the number of species in the category. Suppose that the distribution of the random number of genera can be considered as a *sum* of any number of independent components. Then it would seem possible to consider the numbers of species in the subgenera as random variables adding to the observed number of species. This can be done with Willis variables because of their approximate stability. The problem is then, conceptually, like that of the partition of the energy of
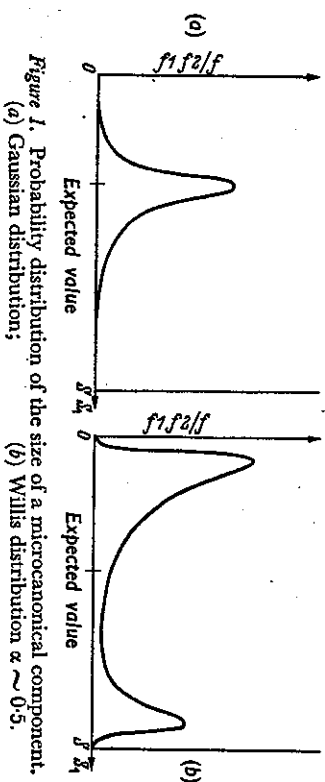

Figure 1. Probability distribution of the size of a microcanonical component. (*a*) Gaussian distribution; (*b*) Willis distribution $\alpha \sim 0.5$.

a large thermodynamical system between two large components; 'energy' is replaced by 'number of species', which changes nothing. It is well known that this partition is 'normally' proportional, with superposed 'fluctuations'; but when the normal distribution is replaced by an exceptional stable one, it follows that there is an overwhelming probability that, if for example, two components are expected to be equal, one is in fact very much the larger. Thus Willis's law formalizes the link between inequality of sizes of genera.

Graphically, let f_1, f_2, f, and s_1, s_2, s, be the probability distribution *a priori*, and the sizes of a component of a genus, of its complement and of the whole. The distribution of s_1 is then, *a posteriori*, if s is known:

$$f_1(s_1)f_2(s-s_1)/f(s) = f_1(s_1)f_2(s-s_1) / \int_0^s f_1(s_1)f_2(s-s_1)\,ds_1$$

If $f(s)$ is the Gaussian, or respectively some Willis distribution, one has the graphs of *Figure 1*, which *clearly show the results quoted.*

First consequence of the preservation of Willis's law in lumping together of independent Willis categories and in splitting of Willis categories

Conjecture that the fact that a species present in area X is, or is not, present in area Y introduces a random splitting of the number of species in X into two independent Willis (α) variables. Then the numbers of species present in *one only* of the two areas, or in *both* simultaneously, are Willis (α) variables; the number of species in both areas is the sum of three independent variables, and the numbers in each area are sums of two variables. Thus Willis's law can be satisfied whichever the area are considered: this is in fact the case, but it was thought to be a proof either of the area or even of the existence of any law. In fact, it is a characteristic stability property of this particular law. Moreover the theory explains the experimental finding

---

that most often one will have one of the following three situations:

(1) very few common species, a large number of local species;
(2) very many common species, a small number of local species;
(3) average number of common species, very many species special to X (resp. Y), very few species special to Y (resp. X).

Second consequence of the preservation of Willis's law in lumping together of independent Willis categories and in splitting of Willis categories

Small changes *upwards*, or *downwards* of the definition of the genera result in lumpings or splittings involving small numbers of Willis categories. Therefore, the Willis property of a system is preserved, another kind of stability.
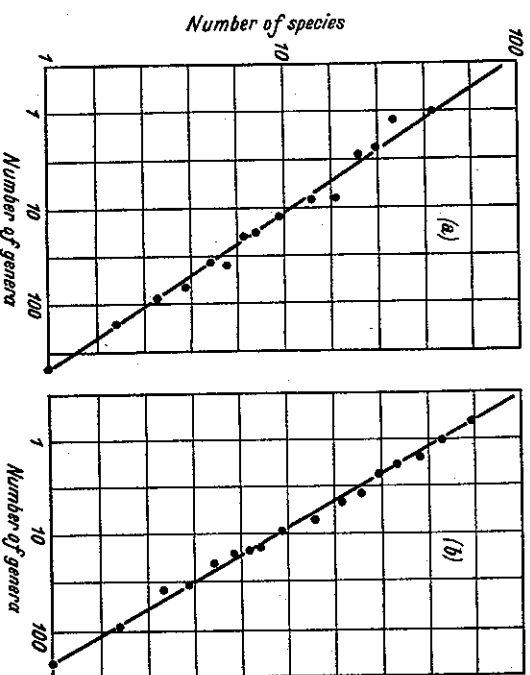

Figure 2. Double logarithmic charts for the frequency distributions of sizes of genera. (*a*) Family—Cerambycinae; (*b*) family—Chrysomelidae. (By courtesy of the Royal Society; Yule[8])

Multiplication: Distribution of species within families

The distribution of species within families can be considered as the 'product' of two distributions: that of genera within families (having the generating function $G_1(u)$) and that of species within genera (having the generating function $G_2(u)$). Then (Feller[8]) the generating function of the 'product' is $G(u) = G_1(G_2(u))$. If both terms are Modified Willis, $G(u) = 1 - \alpha^\alpha(1 - [1 - \alpha^{\alpha_2}(1 - u)]^{\alpha_1})^{\alpha_1} = 1 - \alpha^\alpha(1 - u)^{\alpha_1\alpha_2}$; that is, the product is Modified Willis ($\alpha_1\alpha_2$). This is a kind of homogeneity property. The relationship between S and G can be iterated. The expected number of families is $F = R(\alpha_1)G^{\alpha_1} = R(\alpha_1\alpha_2)S^{\alpha_1\alpha_2}$ etc. Consider now N equidistant categories; all α_i are identical. The order of magnitude of the number of categories, N, required to exhaust S species, will then be such that

$$R(\alpha^N)S\alpha^N = R$$

where R is the number of 'reigns', so that $N \sim \log\log R$.

Generation of the modified Willis (½) systems

Consider the trees, outcomes of a birth and death process with equal rates; that is, such that at each dichotomization there are equal probabilities for the number of species to be one or more than one. The probability that the tree has s species is $2 \cdot 4^{-s}$, the number of different trees with s species is

$$S(s) = \binom{2s-2}{s-1} 2^{-s+1} \sim 4^s S^{-3/2};$$

the probability that a genus has s species is the term in u^s of the development of $1 - (1-u)^{\frac{1}{2}}$; that is, it is given by the modified Willis (½) distribution.

The average (over all trees with s ends) of the average (over the species of a tree) of the number of dichotomies required for identification of equiprobable species is $\sqrt{s}$ (instead of $\log_2 s$ in identification through dichotomies into equal parts). Therefore, the average redundancy tends to 1 as $s \to \infty$.

WILLIS SYSTEMS AS EXCEPTIONAL THERMODYNAMICAL SYSTEMS

Any macroscopic study of systems of a large number of identical elements hinges on some limit theorem of probability. In the case of thermodynamics of matter, or of radiation, some special conditions insure convergence to the 'normal' probability distribution (see KHINCHIN[9]). The framework of thermodynamics is directed towards this application, but is not limited by the special conditions leading to 'normality', and one would think that it can still be used in the study of non-trivial extensions, where geometrical peculiarities of the configuration space lead out of the domain of attraction of the 'normal' distribution. Paradoxically, however, known (and all unsuccessful) generalizations of thermodynamics were not based upon a weakening of any *prior* assumption, but on attempts to re-interpret some *final* results and concepts of the normal case. We shall proceed otherwise, and show that, because of deep geometric peculiarities, the study of taxonomy can be brought into thermodynamics, but only at the early stage. The generalization will be brought at the point where a fundamental limit procedure leads to replace the normal Gauss distribution by an exceptional Cauchy-Paul Lévy distribution. This will preserve 'this fascinating feature of thermodynamics, that quantities and functions, introduced primarily as mathematical devices, almost invariably acquire a fundamental physical meaning' (SCHRÖDINGER[10]). The mathematics of the generalization recalls what is found in gas theory at another stage: order-disorder problems (see MAYER-MAYER[11]).

Some fundamental steps in the Khinchin approach

Let E be the energy of a physical system; let $V(E)$ be the volume of the phase space, included inside the surface of the energy E. The derivative $S(E) = V'(E)$ is called the 'structure function' of the system, and is a kind of measure of the 'number of states' of energy E. Let us sketch Khinchin's approach to the foundations of thermodynamics from the point where it is established that, when the energy E of a whole is given, the probability that a component contains energy E_1 has the distribution $S_1(E_1)S_2(E - E_1)/S(E)$, where $S_1(E)$, $S_2(E)$ are the structure functions of the component studied and of its complement, and where one clearly has $S(E) = \int_0^E S_1(E_1)S_2(E - E_1) \, dE_1$.

The point of a statistical theory of physical objects is to derive properties of bulk bodies without too specific hypotheses on the elementary bodies. Here, one wishes to study the $S(E)$ of sums of a large number of components, without having to assume much about the $S_i(E)$ of each. The symbolic calculus suggests forming the Laplace 'structure generating function' (which becomes identical to the $G(u)$ defined above, if $u = \varepsilon^{-\beta}$)

$$G(\beta) = \int \varepsilon^{-\beta E} S(E) \, dE$$

Then, the $G(\beta)$ of the sum of two systems with g.f. $G_1(\beta)$ and $G_2(\beta)$ is

$$G(\beta) = G_1(\beta) G_2(\beta)$$

For a large number N of components having the same structure generating function $G_1(\beta)$, $G(\beta) = [G_1(\beta)]^N$. To enable one to use the classical theorems of probability, on limits of sums of independent random variables, one arbitrarily forms the 'conjugate probability distributions'

$$p(\beta, E) = \varepsilon^{-\beta E} S(E) G^{-1}(\beta)$$

These distributions are precisely those also reached by Boltzmann's method of the most probable state, where $S(E)$ is introduced as a weighting factor. Here, however, $p(\beta, E)$ need not be the distribution of E in any specified condition. It is simply S multiplied by $G(\beta)$ and by exp $(-\beta E)$, which both cancel out of $S_1 S_2/S$, which is the only expression that matters.

The continuation of Khinchin's argument relies upon 'the fact that $S(E)$ is usually an analytic function which does not increase faster than a certain power of E when $E \to \infty$'. This is the crux of the matter. When it is so, one can roughly expand $\varepsilon^{-\beta E}S(E)$ around its mean value, which is finite, as well as its variance; there is a compromise between the increase of $S(E)$ and the decrease of exp $(-\beta E)$. Most rigorously Khinchin applies a (local) law of large numbers to his 'conjugate distribution'. In all the 'normal' cases, one obtains Khinchin's form

$$S(E) = G(\beta) \exp \beta E \left[\frac{1}{\sqrt{2\pi B}} \exp \left\{ - \frac{(E-A)^2}{2B} \right\} + \text{small order term} \right]$$

The validity of statistical thermodynamics hinges therefore on the conjecture, which could be made a part of 'atomic theory', that the hypotheses of this purely mathematical limit theorem are satisfied by matter.

Generalization of Khinchin's approach to fast increasing $S(E)$

Suppose however that we dismiss any special assumptions. Then, in the general case, $G(\beta)$ is finite only if β is greater than some 'critical value' β^*, the abscissa of (all kinds of) convergence of the Laplace transform $G(\beta)$.

Let $\beta_0 = \sup(\beta^*, 0)$ (that is: $\beta_0 = \beta^*$, if $\beta^* > 0$; $\beta_0 = 0$, if $\beta^* < 0$). Both 'entropy', $H = -\int p(\beta, E) \log p(\beta, E) \, dE$, and 'average energy', $\bar{E} = \int E p(\beta, E) \, dE$ are finite and increasing if $\beta > \beta_0$, and $= \infty$ if $0 < \beta < \beta_0$. The quantity $G(\beta_0)$ may be infinite e.g. when $S(E) = \exp(kE)$. This is found in the case of the number of words of a given cost E, with 'natural' segmentation of languages into words (see MANDELBROT[12,13]).

Let now $G(\beta_0) < \infty$. Boltzmann's approach: maximization of entropy H

at a given average energy $\bar{E}$ may then be continued thus: after the maximization of H at given $\bar{E}$ is performed, maximize relative to all values of $\bar{E}$. This leads one to choose for $1/\beta$ precisely the 'critical temperature' $1/\beta_0$. However to this temperature corresponds now a proper probability distribution of E; whereas in the normal case, or if $G(\beta) = \infty$, the maximization of H, without fixing $\bar{E}$, leads to the improper distribution for which all states are empty.

For example, it was seen in the generation of Willis (✠) systems that the number of configurations $S(s)$ of given energy (number of species) increases like $4^s s^{-3/2}$, so that Boltzmann's formula $p(x) = S(x) \exp(-\beta x) G^{-1}(\beta)$ can be defined for $\beta > \log 4$ only. At the limit temperature one has Willis (✠) systems, and the sum of a large number of such systems does not follow the usual normal distribution, but a stable distribution of index $\frac{1}{2}$.

Altogether, one part of Khinchin's hypotheses amounted to $\beta_0 = 0$; even if this requirement is dropped most of Khinchin's argument remains valid when $\beta > \beta_0$. (Some special considerations which are brought by the finiteness of β_0 were given by the author in the study of the thermodynamics of the rank frequency relationship for words—MANDELBROT[12,13].)

One cannot have $\beta < \beta_0$, except when a cut-off of E can be introduced, so that the really new phenomenon can occur only if $\beta = \beta_0$. This is the object of the present study, and introduces an apparently new kind of 'thermodynamic fine structure at a given temperature', of a seemingly exceptional type, but which in fact applies to taxonomic categories.

What other behaviour for $S(E)$ could be encountered a priori in natural systems? Thermodynamics assumes that it is adequate to describe a bulk body by a single probability distribution; this can only mean that, if and when the body is large, variations of its size do not change the form of the distribution, but only a few scaling factors. Practically, non-linear scaling cannot be considered in this paper, because few theorems are available about limit tendency of non-linear functions of random variables; but theorems of Paul Lévy and I. A. Khinchin, and of B. V. GNEDENKO[14] solve the linear case. They state that, in order that a distribution function $F(x)$ be a limit distribution for sums $y_n = B_n \Sigma_1^n x_i - A_n$ where the x_i are independent identically distributed addends, it is necessary and sufficient that $F(x)$ be 'stable', that $B_n = n^{-1/\alpha}$, and that the distribution of x_i behave at infinity like $x^{-\alpha}$ (this last hypothesis is much stricter than those sufficient for tendency to the normal).

We have exhibited one family of stable distributions. There is also another family, since our formula for $\varphi(t)$ remains meaningful for $1 < \alpha < 2$; but that is all. Thus, besides Willis systems, the only systems with possible non-normal thermodynamic structure will correspond to stable distributions of index $1 < \alpha < 2$ (which occur in econometry, as is shown elsewhere by the author) for which probability still behaves like $Px^{-(\alpha+1)}$ for x large.

Though no counterpart seems to be in view for Willis structure, the study of systems with Willis structure seems to be in view for a more natural and less trivial use of thermodynamic thinking in information theory than in the study of 'analogies between information and entropy'.

The probability distributions such as $Px^{-(\alpha+1)}$ were believed by ZIPF[3] to be

as central to random phenomena in social sciences as the normal Laplace-Gauss distribution is to random phenomena in physical sciences. This has led him to conjecture the existence of some common principle of 'least effort', a variational principle of behaviour. In fact, the theories of the various Zipf's laws have to be based on very variable assumptions, as to what is the appropriate random population, which makes the existence of a single principle very problematic. However, Zipf's belief in the importance of these probability distributions may very well be vindicated by several distinct theories. One other was given in MANDELBROT[12,13], for the case of the observed word frequencies; still other models are given elsewhere.

MODELS FOR WILLIS SYSTEMS

It is only fit and appropriate that a model for the Willis distribution has the deep homogeneity properties of the law itself. A number of such models are found in the counterparts of a few approaches to 'normal' thermodynamics. They may be classified in various ways:

(a) in any case (and unless one is ready to consider signs and names as part of the outside world), any property of a result of observation may be considered as due, either to the thing observed, to the observer, or to the items taxonomized, to the taxonomist, or the taxonomy itself; in other words is biological, psychological or linguistic.

(b) the models may consider the observed law, either as necessary, or as due to a choice not involving chance, or as involving a random mechanism, without or with the use of a limit theorem. The limit may refer to an actual limiting procedure taken with respect to time: such models are called diachronic, models not involving time being synchronic.

(c) a model can consider the system in bulk (normalized form) or as a composition of elementary systems (extensive form).

Synchronic 'normalized' psychological models involving chance, but no limit theorems

In order to choose at random a whole taxonomic tree for a family of S species, one needs a probability measure on the set of all trees with S endpoints, and with a transversal line of genera. If all such trees are equiprobable the relationship between s and g in the average taxonomy is bound to be Willis (✠). Willis's law therefore expresses 'disorder' of natural taxonomies. This is closer to normal thermodynamics than is the search for order.

One may also use the continued Boltzmann maximization of entropy, and one finds that Willis (✠) is the most probable taxonomy. (This provides a criterion of 'equilibrium' without balancing of forces or definition of potential; or else an actual variational principle of behaviour.)

Other values of α would result from other rules of discernibility for taxonomic trees.

These most probable or average taxonomies also have the property of being very inefficient (redundancy $\to 1$ as $S \to \infty$). This is a distinctive property; any criterion of splitting of genera leading to extremely unequal parts requires genera having a probability distribution behaving like

$e^{-bs}f(s)$, where $f(s)$ decreases slower than any damped exponential. If this property is to hold for the subgenera obtained by a few splittings, b must be zero. Further, suppose now that the distribution $p(s)$ is to remain of the *same type* after any number of splittings; that is, the tree is to be homogeneous, and it is to be impossible to decide, from $p(s)$, at which 'level' of the tree one finds oneself. Then $f(s)$ must be a close approximation to a stable distribution, that is, to Willis's distribution.

An extensive model involving chance, and a limit theorem: The biological diachronic (phylogenetic) model of Yule

G. UDNY YULE[2] considers Willis's law as a law due to the regularity of the random process of evolution. Genera and species are assumed to be fully intrinsic and their evolution to be ruled by chance alone, and to be representable by multiplications (splittings) with constant average rates γ and α. Let $\alpha = \gamma/\sigma$. Then, after a very long time of evolution, one obtains:

$$p(s) = \frac{1}{s}\,\frac{1}{(1+\alpha)\left(1+\frac{\alpha}{2}\right)\cdots\left(1+\frac{\alpha}{n}\right)} \sim \alpha\Gamma(1+\alpha)s^{-(\alpha+1)}$$

which is another modified form of Willis's probability distribution.

An extensive linguistic, diachronic (phylogenetic) model

A system of signs is first established by some taxonomist; but then, it follows a sort of 'evolution': as new species are found, other, earlier considered as different, are found to be the same, new areas are discovered etc. In the spirit of the theory of communication, Yule's theory could be re-phrased by denying all reality to either species or genera, but imagining that the words that represent them follow an evolution, largely independent from the point of departure. This may be considered as a reasonable theory for trade names and business catalogue indexes where, clearly, an evolution has proceeded within the history of language; but it is not so in the case of biological taxonomies, often established by a single man, once for all.

Synchronic limit models

If Willis's law were established for subcategories adding to a category (at least approximately established), it would follow accurately for the total category. However, such applications of limit theorems are less satisfactory here than in gas theory, since observed samples are small in numbers, whereas the 'elementary cells' are to have infinite expected number of items, which are contradictory conditions. (Note that in quantum statistics elementary cells are also so small that the method of most probable state of Boltzmann fails.) The limit procedure could also be applied to the iteration of the multiplication of an increasing number of distributions with α close to 1.

Observational models

(a) Imagine that the stability, homogeneity *etc* properties of Willis's systems are *required* of satisfactory taxonomies, and even preferred to small redundancy. After the taxonomist has considered many alternatives he would necessarily (but unconsciously) choose a Willis system.

(b) A system is left alone when no further small change can 'improve' it; that is, change its statistics appreciably.

(c) All observed category systems have the Willis structure, because no other structure *could have been noticed*, to start with, under the conditions of observation (local and global floras *etc*). Stable distributions are the only ones, for which no difficulty can come from such remarks as 'we seek to measure what we want to measure; we often end by measuring what we can'.

(d) If one had to form subjective *a priori* probability distributions, a principle of sufficient reason would lead to stable distributions (independence from the area investigated *etc*).

Necessary models

These could be diachronic or synchronic. As an example of the first case, imagine an evolutionary process not involving chance; it applies to the things taxonomized. As an example of the second case, imagine strict 'logical' relationships between the items of the taxonomy. We do not believe that Willis's law could be derived in this way. However, in any case, 'local' biological or linguistic laws are bound to be relevant to bulk properties of the taxonomy, and there is a very real problem in explaining how they may coexist with the Willis 'bulk' properties: a problem fully analogous to the ergodic problem in physics.

REFERENCES

1 WILLIS, J. C. *Age and Area*, Cambridge; University Press, 1922
2 YULE, G. UDNY. *Phil. Trans.*, B, 213 (1924) 21
3 ZIPF, G. K. *Human Behavior*, Cambridge (Mass.); Addison Wesley, 1949
4 GOOD, I. J. *Biometrika*, 40 (1953) 237
5 FELLER, W. *Trans. Amer. Math. Soc.*, 67 (1949) 98
6 LÉVY, PAUL. *Calcul des Probabilités*, Paris; Gauthier-Villars, 1925
7 — *Addition de variables aléatoires*, Paris; Gauthier-Villars, 1937
8 FELLER, W. *Probability Theory*, Section 11.5, New York; J. Wiley, 1950
9 KHINCHIN, I. A. *Statistical Mechanics*, New York; Dover, 1949
10 SCHRÖDINGER, E. *Statistical Thermodynamics*, Cambridge; University Press, 1949
11 MAYER, J. and MAYER, M. G. *Statistical Mechanics*, Chaps. 13–14, New York; McGraw Hill, 1940
12 MANDELBROT, B. *Communication Theory*, P. 486, London; Butterworth, 1953
13 — *Diagnostic en l'absence de bruit*, pp. 1–90, Institut de Statistique de l'Université de Paris, 1955
14 GNEDENKO, B. V. and KOLMOGOROV, A. N. *Limit theorems for sums of random variables*, Cambridge (Mass.); Addison Wesley, 1953

NOTE ADDED IN PROOF. A variant of Yule's model for Willis's distribution has recently been given in *Biometrika* 42 (1955) 425 by H. A. Simon (who credits the distribution to Yule). However that author nowhere notes *the approximate stability of the distribution, which appears to us to be the crux of the matter, irrespective of any* model.