Progress in Biocybernetics

N. Wiener
J. P. Schadé

PROGRESS IN BIOCYBERNETICS

VOLUME 2

EDITED BY

NORBERT WIENER

Massachusetts Institute of Technology, Cambridge, Mass. (U.S.A.)

AND

J. P. SCHADÉ

Central Institute for Brain Research, Amsterdam (The Netherlands)

ELSEVIER PUBLISHING COMPANY

AMSTERDAM — LONDON — NEW YORK

1965

ELSEVIER PUBLISHING COMPANY
335 JAN VAN GALENSTRAAT, P.O. BOX 211, AMSTERDAM

AMERICAN ELSEVIER PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE, NEW YORK, N.Y. 10017

ELSEVIER PUBLISHING COMPANY LIMITED
RIPPLESIDE COMMERCIAL ESTATE, BARKING, ESSEX

LIBRARY OF CONGRESS CATALOG CARD NUMBER 64-18530

WITH 71 ILLUSTRATIONS AND 2 TABLES

PRINTED IN THE NETHERLANDS

LIST OF CONTRIBUTORS

B. L. M. CHAPMAN, Educational and Scientific Developments Ltd., Bristol (Great Britain).

D. GABOR, Department of Applied Electron Physics, Imperial College University of London, London (Great Britain).

O. H. IVERSEN, University of Oslo, Department of General and Experimental Pathology, Oslo (Norway).

C. A. MUSES, International Research Centre for Mathematics, Morphology, and Morphogenesis, Lausanne (Switzerland).

A. NEWELL, Carnegie Institute of Technology, Pittsburgh, Pa. (U.S.A.).

A. RAPOPORT, The University of Michigan, Department of Psychiatry, Ann Arbor, Mich. (U.S.A.).

N. RASHEVSKY, Mental Health Research Institute, The University of Michigan, Ann Arbor, Mich. (U.S.A.).

W. ROSS ASHBY, University of Illinois, College of Engineering, Urbana, Ill. (U.S.A.).

J. SAUVAN, 43, Rue Lacépéde, Paris Vème (France).

H. A. SIMON, Carnegie Institute of Technology, Pittsburgh, Pa. (U.S.A.).

G. VOSSIUS, Institut für Animalische Physiologie der J. W. Goethe Universität, Frankfurt a.M.-Süd (D.B.R.).

PREFACE

Early in 1964 Prof. Norbert Wiener came to Amsterdam to act as visiting professor in cybernetics at the Central Institute for Brain Research.

During the first month of his stay we discussed in great detail the publication of a second and a third volume in the series Progress in Biocybernetics. In the course of the conversation with some senior members of our staff we decided to publish the second volume of the series as a celebration volume, marking the occasion of Prof. Wiener's 70th birthday in November 1964. We decided to let him choose the contributors for two celebration volumes (volume 17 of our series Progress in Brain Research was also meant to be a Festschrift for Norbert Wiener, containing in particular papers related to the nervous system) without letting him know that the contributors were asked to write a survey of their work in honour to the great pioneer of cybernetics.

Due to the untimely death of Norbert Wiener in March 1964 we had to change the celebration volume into a memorial volume.

In this volume a series of contributions are assembled which give examples of the wide application of cybernetics to various areas of biology and medicine. The papers are either original contributions, or reviews of significant applications of cybernetics to various areas of the biological sciences. The scientists are all outstanding in their field and have in the past made important contributions to biocybernetics.

J. P. SCHADÉ

CONTENTS

THE SMOOTHING AND FILTERING OF TWO-DIMENSIONAL IMAGES

DENNIS GABOR

Department of Applied Electron Physics, Imperial College, University of London, London (Great Britain)

It is meet and right on this occasion to turn one's thoughts to Norbert Wiener's monumental early work on the Extrapolation, Interpolation and Smoothing of Stationary Times Series (1949) and to consider a line which branches off from it, as a natural extension, but which so far has received far too little attention.

Two-dimensional images suffer from distortion and noise just as much as one-dimensional messages. In television a major part of the problem is one-dimensional. A re-distortion method for television signals was demonstrated by Goldmark and Hollywood in 1951, under the name of 'crispening'. This was extended to two-dimensional images by Kovasznay and Joseph in 1955, who showed that complete re-distortion is possible in principle if each picture point is blurred by a diffusion process. The first term in their correcting series is the Laplacian of the brightness, and to this point the correction could be realised experimentally, with impressive results. However, quite apart from instrumental difficulties, the Kovasznay–Joseph series could not be continued indefinitely because it would soon start to amplify picture noise to an intolerable extent.

I am not aware of attempts to tackle picture noise in two dimensions, though the problem is of immense practical importance. Geneticists are very worried about the possible cumulative effects of radiography, and they insist on a minimum of radiograms, especially in pregnancy, with minimal doses of X-rays. This leads logically to radiograms with as few photons as are required to convey the information. Professor W. V. Mayneord, who first drew my attention to this problem, has constructed a scanning device, in which hardly a single X-ray photon need be wasted, but which naturally gives very noisy pictures. Similar problems, though of less urgency, arise in the electron microscopy of radiation-sensitive objects, and also in photo-

References p. 9

graphy and television transmission under conditions of very restricted illumination. It therefore appears worthwhile to consider the problem of smoothing such noisy pictures with a maximum extraction of information.

So little appears to have been done in this field, that I feel justified in attacking it here in a somewhat light-handed way, disregarding all mathematical rigour, and not taking 'optimum' solutions too seriously. I could hardly do otherwise, because from the start we lack the solid foundation

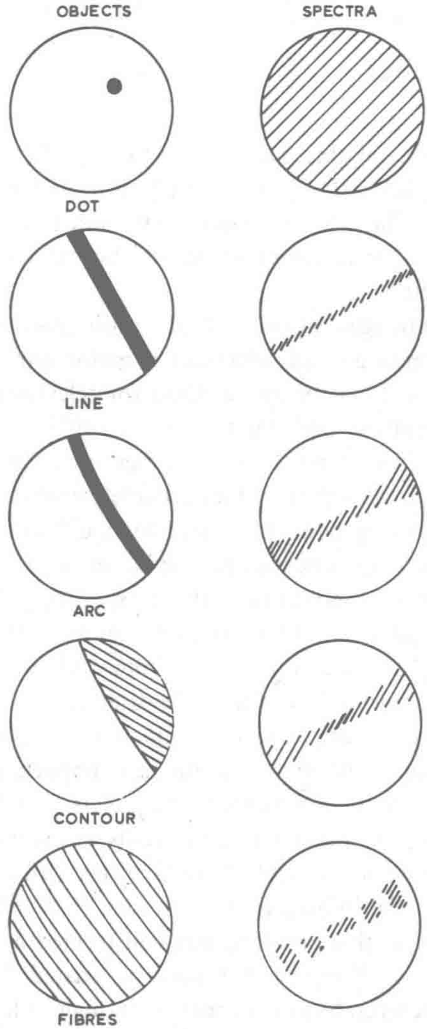

Fig. 1. Image elements and their two-dimensional Fourier spectra.

of the Wiener–Shannon communication theory: the probability pattern of messages. The situation is somewhat similar to the one in the processing of speech, in which little headway has been made in fifteen years, because we have not yet succeeded in extracting the characteristics of meaningful sounds from the enormous variety of possible sounds. Superficially it appears even worse, because an interval with 100 Nyquist samples, each capable of 10 distinguishable levels has 10^{100} possible configurations, but this is comparable to a picture area containing 100×100 independent elements with 10 levels each, which has $10^{10\,000}$ possible states. In reality the situation is not quite so bad. The reasons for my optimism are explained in Fig. 1. At the left-hand side I have shown certain shapes in the image plane x, y in a small field which may be called the processing area. At the right-hand side I have shown the absolute values of the amplitude in their Fourier transforms, in a Fourier plane ξ, η. If the image is a transparency in a microscope and is illuminated with plane monochromatic waves parallel to the optic axis, the Fourier transform appears physically as the light amplitude in the rear focal plane of the objective. For a start the Fourier plane may be limited only by a circular aperture corresponding to the maximum spatial frequency which the optical system can transmit.

Significant shapes are very rarely composed of independent dots. I will leave this rare case out of account, and consider rather the far more frequent shapes shown below the first; a straight line, an arc, a contour and a bundle of (not striated!) fibres. If we construct the two-dimensional Fourier spectra of these shapes, we see that all four are restricted to a small fraction of the Fourier plane; to a more or less narrow strip. This strip always passes through the centre of the Fourier plane, corresponding to zero spatial frequencies, but its direction varies. This at once shows up a significant difference between the Wiener–Kolmogoroff one-dimensional filtering theory and the two-dimensional case: Even the most primitive two-dimensional filter must be *adaptive*. It must adapt itself to the principal direction in that part of the image which we take as the unit of processing.

The choice of this processing area is of course a very important one for the success of the operation. It must be significantly larger than the resolution area; large enough to recognize simple shapes such as those shown, but not so large as to admit more complicated ones except in rare cases, and it must of course contain a sufficiently large number of elementary pulses, (which we will simply call 'photons') to make recognition of the shapes possible. In an ideal machine the size of this area would itself be adaptable; small where the photon density is large, and *vice versa*. The human eye

References p. 9

undoubtedly makes use of such adaptive elementary areas, but it does much more than this. We cannot follow it here in all its perfection. In the present analysis we can consider the elementary processing area as a constant, but we compensate to some extent for this imperfection, because we use the elementary area for obtaining only one datum; the most probable density at its centre.

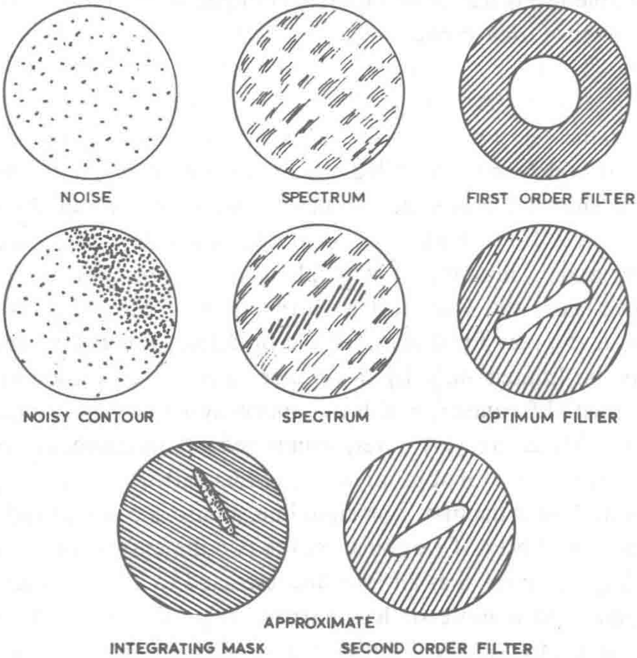

Fig. 2. Smoothing and filtering a noisy image.

In Fig. 2 we face the problem of extracting this datum from the noise. Strictly speaking this is the fundamental problem of statistics; 'what is the probability distribution of which the given image is a sample?'. But a statistical approach is unavailing; the number of hypotheses to be tested is too overwhelming. I prefer a step-by-step approach which starts with 'first order smoothing'. This is exactly the same as what the eye does when presented with a noise image. We either step back, until the noise dots do not appear separate, or we look at it through half-closed eyelids.

Mathematically this means that we replace a noise dot at x_0, y_0 by some smoothing unit function $u(x - x_0, y - y_0)$, for instance by the Gaussian function

$$u = (k^2n/\pi) \exp \{-k^2n [(x-x_0)^2 + (y-y_0)^2]\} \quad (1)$$

where n is the number of photons per unit area and k is a numerical factor of the order unity. This could be determined for instance in such a way that if the n dots are distributed in a regular hexagonal pattern the minimum between them just vanishes. Selecting the optimum smoothing function is of course quite a study in itself, but we will not pursue it further, because it is only a first step in our process. In the first row in Fig. 2 I have shown a random pattern of dots and indicated its spectrum and the 'first order aperture'. If the unit smoothing function is Gaussian, the spectral aperture is also Gaussian; it has been shown in Fig. 2 as sharp only for simplicity.

The second row in Fig. 2 shows an example of a noisy image and its Fourier spectrum. It is seen that something has been achieved by considering the Fourier transform instead of the image. While the image itself can be said to be 'composed of noise', in the Fourier representation this has become 'background noise'. With our visual intelligence we could draw a contour around the significant part of the spectrum, and thereby exclude most of the noise. In principle a machine could do this too, for instance the contour enhancer of Kovasznay and Joseph, if it is given certain instructions. But for the present I do not want to aim too high — a machine capable of such a high degree of intelligent discrimination might defeat itself by its complexity — and propose a simpler process.

The next stage in this process is to reprint the original noisy image with its smoothing function, that is to say replace it by the amplitude

$$f(x,y) = \sum_{m}^{N} u(x-x_m, y-y_m) \quad (2)$$

summed over all photons in the processing area. It is advantageous to give the processing area washed-out instead of sharp limits, for instance by limiting this too by a Gaussian aperture, as a sharp image field gives a system of circular fringes in the Fourier plane, which present an unnecessary complication. It is also advantageous to consider not the brightness, (or blackness,) of the image as the 'amplitude', but its square root, and in what follows, we will use the term in this sense.

Though the smoothing has reduced the noise in the picture, and has contracted the area in the Fourier plane, it has not made the recognition of the contours of the significant area in the spectrum any easier. I propose therefore a less ambitious method, which at least has the advantage that it leads to not too complicated instructions.

I propose to select the significant area in the Fourier plane by determining

References p. 9

the mean square frequency as a function of the direction. For simplicity I write down the definition and certain alternative expressions of the mean square frequency in one direction only. Let $f(x)$ be the amplitude and $F(\xi)$ its Fourier transform

$$f(x) = \int_{-\infty}^{\infty} F(\xi)\, e^{2\pi i x \xi}\, d\xi \qquad F(\xi) = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i x \xi}\, dx \tag{3}$$

then the mean square frequency $\overline{\xi^2}$ is (Gabor, 1946),

$$\overline{\xi^2} = \frac{\int_{-\infty}^{\infty} F^* \xi^{\ 2} F d\xi}{\int_{-\infty}^{\infty} F^* F d\xi} = -\frac{1}{(2\pi)^2} \frac{\int_{-\infty}^{\infty} f^* \frac{d^2 f}{dx^2} dx}{\int_{-\infty}^{\infty} f^* f dx} = \frac{1}{(2\pi)^2} \frac{\int_{-\infty}^{\infty} \left(\frac{df}{dx}\right)^2 dx}{\int_{-\infty}^{\infty} f^2\, dx} \tag{4}$$

The last transformation is valid if f or df/dx vanish at infinity, or (as in our case) at some finite boundaries. The denominator is the total brightness (or blackness) in the interval which is also a measure of the total pulse number in it. Thus the mean square frequency is essentially the mean square gradient of the amplitude. (This is similar to, but not quite the same as, what Cherry and Gouriet (1953) have called 'picture detail' in a television line, because they have used the brightness signal instead of the amplitude, which is its square root.)

In two dimensions the mean square frequency $\overline{\varrho^2}$ in a direction θ

$$\varrho = \xi \cos\theta + \eta \sin\theta$$

is conveniently expressed in the form

$$\overline{\varrho^2} = \frac{1}{(2\pi)^2 \int_{-\infty}^{\infty} f^2 dx dy} \left\{ \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x}\right)^2 dx dy \right] \cos^2\theta + \right.$$

$$\left. + \left[2 \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial x} \frac{\partial f}{\partial y}\right) dx dy \right] \sin\theta \,.\, \cos\theta + \left[\int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial y}\right)^2 dx dy \right] \sin^2\theta \right\} \tag{5}$$

This has the form

$$\overline{\varrho^2} = A \cos^2\theta + B \sin^2\theta + 2C \sin\theta \,.\, \cos\theta \tag{6}$$

If we represent the root mean square frequency as a radius vector in the ξ, η plane this is a curve of the fourth order, which is an approximation to the optimum mask in the Fourier plane. It is more convenient to place the mask into the object plane, because if for $\overline{\varrho^2}$ we substitute its reciprocal $r^2 = 1/\overline{\varrho^2}$ we obtain an ellipse

$$Ax^2 + By^2 + 2C\, xy = \text{constant} \tag{7}$$

I propose therefore to take as the second smoothing filter a Gaussian mask, with the transmission

$$\exp \{-\pi^2 [A(x-x_0)^2 + B(y-y_0)^2 + 2C(x-x_0)(y-y_0)]\} \tag{8}$$

which is so determined that in any direction θ the root mean square frequency is transmitted with 1/e of the transmission for zero frequency. As shown in Fig. 2, this mask is narrow at right angles to the contour, where we want to transmit high spatial frequencies, and long parallel to it, where the frequencies are low.

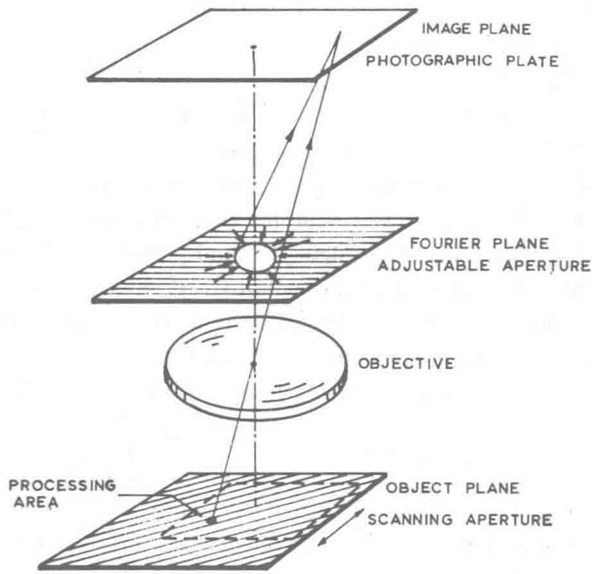

Fig. 3. Realisation of the first smoothing by a photographic process.

It remains now to show that this smoothing process can be carried out with not too complicated instrumentation. For the first smoothing I propose an optical process, as illustrated in Fig. 3. The starting point is a transparency which is introduced in the object plane of an optical system such as a microscope, and is illuminated with monochromatic light, parallel to the axis. This is scanned with an aperture the size of the 'processing area'. The Fourier plane, that is to say the rear focal plane of the objective, contains an adjustable aperture centered on the axis. By eqn. (1) the area of this aperture is made proportional to the density of photons in the processing area, up to the maximum aperture of the optical system. This can be achieved by

References p. 9

servoing the aperture from a photometer, which measures the transmitted light. The aperture can be made approximately 'Gaussian' by the simple trick of arranging it a little off the Fourier plane. In this way we obtain a photograph, which is smoothed to the extent that the 'photons' do not appear any more as separate dots.

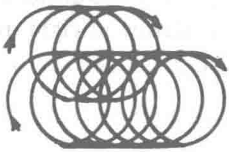

Fig. 4. Circling scan.

The next step is determining the coefficients A, B and C in eqn. (8) from eqn. (5). I propose projecting the transparency on the face of an image camera, which is electronically scanned with a finely focused beam, in the way as illustrated in Fig. 4. The scanning point rotates in a small circle of radius r with a high angular frequency ω, and progresses slowly in the line direction. The process is simplest if the photograph was taken with a 'gamma' of 0.5, as in this case the transmitted light is proportional to the square root of the original brightness, but the square root can be also produced electronically. The quantity read off by the electron beam is then

$$F = f(x,y) + r \left(\frac{\partial f}{\partial x} \cos \omega t + \frac{\partial f}{\partial y} \sin \omega t \right) \tag{9}$$

and this is processed as follows:

Filtering out the high frequency ω gives $f(x,y)$. This is squared and integrated in the processing area, giving the denominator in eqn. (5).

Squaring F and filtering out the high frequencies ω and 2ω gives

$$f^2 + \tfrac{1}{2}r^2 \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right] \tag{10}$$

Multiplying the square of F by cos $2\omega t$ and filtering afterwards gives

$$\tfrac{1}{4}r^2 \left[\left(\frac{\partial f}{\partial x} \right)^2 - \left(\frac{\partial f}{\partial y} \right)^2 \right] \tag{11}$$

Multiplying by sin $2\omega t$ and filtering gives

$$\tfrac{1}{2}r^2 \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} \tag{12}$$

From the three quantities 10, 11 and 12 a computor coupled to the image camera can compute the integrands in the numerator of eqn. (5) and, after integration, by division with the total brightness also the quantities A, B, C It can then, if it has stored the values $f(x,y)$ in the processing area, calculate the most probable value of $f(x_0, y_0)$ by multiplication with the mask function 8 and a final integration over the processing area. This is fed into a display device, for instance a cathode-ray tube, which displays the twice processed image.

Thus, by a combination of optical devices, electron cameras and fast computors the smoothing process which I have proposed can be carried out. It is also obvious that nothing less than saving humanity from degeneration by radiography could ever justify a process of such complication, — except perhaps military interests! For the moment considerations such as those sketched out here have at any rate one merit. They make us feel due respect for the brain plus its associated computor which carries out in a flash such operations, and a lot more besides.

SUMMARY

In the great majority of two-dimensional images the objects of interest are not points, but contours. This feature enables us to lift them out of a noisy background (photon noise or photographic grain) by a two-stage process. In the first stage, the noise is smoothed by a 'first order spectral filter', so adjusted that the noise dots optimally merge. In the second stage the direction of the contour is recognised and an anisotropic spectral filter is applied such that the definition is poor in the direction of the contour, and as good as possible at right angles to it. A scanning method operating with a cycloidal scan is proposed, and a mathematical process is described which allows computing the second filter and correcting the image in a continuous operation.

REFERENCES

Cherry, E. C., and Gouriet, G. G., (1953); Some possibilities for the compression of television signals by recoding. *J. Inst. Elec. Eng.*, **100**, 9–18.

Gabor, D., (1946); Theory of communication. *J. Inst. Elec. Eng.*, **93**, 429–457.

Goldmark, P. C., and Hollywood, J. M., (1951); A new technique for improving the sharpness of television pictures. *Proc. IRE*, **39**, 1314–1322.

Kovasznay, L. S. G., and Joseph, H. M., (1955); Image processing. *Proc. IRE*, **43**, 560–570.

Wiener, N., (1949); *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. New York. Wiley.

CONSTRAINT ANALYSIS OF MANY-DIMENSIONAL RELATIONS

W. ROSS ASHBY

University of Illinois, College of Engineering, Urbana, Ill. (U.S.A.)

That this article should be offered as a tribute to Norbert Wiener is specially appropriate, for it takes as basis an observation of his that has not yet shown, I feel sure, its full fertility. I refer to his original suggestion (Wiener, 1914) that a 'relation', previously regarded as somewhat metaphysical, be identified, at least for operational purposes, with the set of those n-tuples that satisfy the relation. At one stroke the 'relation' becomes an ordinary mathematical object that can be subjected to the ordinary mathematical operations, even when the relation is wholly arbitrary.

To the biologist, the freedom that allows it to be wholly arbitrary is most welcome, for in his science, though relations are of the greatest importance, they seldom have the tidiness common to more formal mathematics.

The attempt to apply Relation Theory to the biological sciences soon runs into great difficulties however. As soon as the biologist attempts to deal with realistically large numbers (*e.g.* with 10^{10} nerve cells) or with realistically intricate patterns of interaction (*e.g.* those between species in the Amazon jungle) the combinatorial possibilities soon generate fantastically large numbers; equally fantastic is the quantity of information-processing demanded of any system (cerebral or electronic) that would handle the questions involved. Exponentials, factorials, or even more explosively increasing functions appear. Bremermann (1962) has shown that no computer made of matter, and therefore subject to the mass-energy relation and to Heisenbergian uncertainty, can possibly process more than 1.4×10^{47} bits per g per sec; so 10^{70}, say, is certainly an absolute upper bound to what is practical. Yet even the simplest problems with more than a few variables and more than infinitesimal interaction generate numbers vastly greater than 10^{70}. An example is given below. Any science, such as cybernetics, that would treat large systems with strong interactions, urgently needs methods by which the excessively complex can be reduced to complexities that are within our resources. In this paper is described one such method.

The method is based on the common observation that when the number of variables is large — a thousand and over, say — many of the significant relations are not really intricate to the full degree, but are really built out of simpler relations. In dynamics, for instance, the linear system, both common and important, has the peculiarity that the complicated output evoked by a complicated input can be found by simply adding a number of simple outputs, each evoked by a simple input. Thus in this case the whole output–input relation is really composed of many simple relations that combine only by adding. The parts of the system interact, but not the sub-relations.

Again, a camera lens with ten elements and fifteen surfaces at first seems optically very complex; yet in fact the total effect, from incident ray to emergent ray, can be obtained by merely repeating one *ternary* relation (incident ray/surface/refracted ray) fifteen times in succession. Thus the lens designer is able to avoid the fantastic combinatorial possibilities initially presented by the 15-variable relation.

Not only the physical sciences but everyday life shows the same feature. 'The Law', as it affects John Citizen, has hundreds, even thousands, of variables. Yet it can, in fact, be dealt with piecemeal; for it is built by the intersection of such sub-relations as: Drivers of age x may drive only automobiles of class y; Stores selling goods p must be closed on days q.

The set of all events that are 'legal' is then obtained by simple compounding of all the sub-relations, each of which uses only a tiny fraction of the totality of variables.

An indication of what threatens may emphasize the point here. If there are n variables, and each variable can take k values, the number of relations possible (*e.g.* of Laws that J. C. may face), as subsets of the product set, is

$$2^{(k^n)}.$$

As a function of n, it is an exponential of an exponential, a rate of increase vastly more 'explosive' than those commonly encountered in other branches of science. If k is merely 10, for instance, by the time n has risen to five (well below the 15 of the 'optical' example) the number of relations has risen to about $10^{30\,000}$, a number that shows how intensely restrictive Bremermann's limit really is. Let us then consider how one complex n-ary relation may be reduced to a set of simpler relations.

When the dimensions are 2, the relation, as a subset of a product-set $E \times F$, has only the simplifying possibility that it is itself a product set,

References p. 18