ranging, by both procedures, from 29.0 to 30.8 percent. The slight discrepancy with the results of Wyatt and Cohen (*19*), who reported 17.1 percent G and 15.4 percent C in *R. prowazeki*, can be attributed to differences in methods. The spotted fever group, represented by two strains of *R. rickettsi* and one strain each of *R. conori* and *R. akari*, the etiological agents of Rocky Mountain spotted fever, boutonneuse fever, and rickettsialpox, respectively, also have G + C contents indistinguishable from each other with a range of 32.0 to 33.3 percent. The discrepancy between this percentage and that obtained by Price (*20*), who reported G or C contents ranging from 17.3 to 19.4 for *R. rickettsi*, cannot be explained. *R. canada* falls within the range of the typhus group rickettsiae. The trench fever rickettsia, *R. quintana*, has a considerably higher G + C content (38.6 percent).

The following conclusions can be cautiously drawn from the results described above. *R. quintana*, despite its similarity to *R. prowazeki* in ecological association, which involves man and the body louse, has only a remote evolutionary relationship to this organism. It is entirely justified to consider that the biologic characteristics of *R. quintana* which distinguish it from other rickettsiae, that is, extracellular growth in the gut of the louse and in bacteriological media (*21*), are reflections of profound genetic differences. On the other hand, we are justified also in continuing to regard epidemic and murine typhus rickettsiae as closely related organisms. The G + C content of approximately 30 percent for the typhus group and approximately 32.5 percent for the spotted fever group indicates an early evolutionary divergence of these organisms from their common ancestor. In contrast, the three species of spotted fever rickettsiae, despite their wide differences in geographic distribution and arthropod host (*1*), have remarkably similar G + C contents. The Q fever rickettsia, *Coxiella burneti*, not included in this study, has a G + C content of 43 to 45 percent (*3–5*) and it must be considered remotely related to the other rickettsiae.

Of particular interest to the understanding of the evolution of rickettsiae are our findings with *R. canada*. This organism was isolated from a tick in North America, but, surprisingly, several of its phenotypic characteristics link it more closely to the typhus than to the spotted fever group (*22*). Our findings suggest that *R. canada* has evolved from a strain of the typhus group and that it is unlikely that it resulted from a limited number of mutations of a strain of the spotted fever group (*23*).

FRANKLIN J. TYERYAR, JR.
EMILIO WEISS
DAVID B. MILLAR
*Naval Medical Research Institute, Bethesda, Maryland 20014*
F. MARILYN BOZEMAN
*Walter Reed Army Institute of Research, Washington, D.C. 20012*
RICHARD A. ORMSBEE
*National Institute of Allergy and Infectious Diseases, Hamilton, Montana 59840*

### References and Notes

1. E. Weiss and J. W. Moulder, in *Bergey's Manual of Determinative Bacteriology*, R. E. Buchanan and N. E. Gibbons, Eds. (Williams & Wilkins, Baltimore, ed. 8, in press).
2. Z. A. Cohn, F. M. Bozeman, J. M. Campbell, J. W. Humphries, T. K. Sawyer, *J. Exp. Med.* **109**, 271 (1959).
3. J. B. Smith and G. P. Stoker, *Brit. J. Exp. Pathol.* **32**, 433 (1951); S. Schramek, *Acta Virol. Prague* **12**, 18 (1968).
4. J. Marmur and P. Doty, *J. Mol. Biol.* **5**, 109 (1962).
5. C. L. Schildkraut, J. Marmur, P. Doty, *J. Mol. Biol.* **4**, 430 (1962).
6. C. L. Wisseman, Jr., E. B. Jackson, F. E. Hahn, A. C. Ley, J. E. Smadel, *J. Immunol.* **67**, 123 (1951); H. G. Stoenner, D. B. Lackman, E. J. Bell, *J. Infect. Dis.* **110**, 121 (1962).
7. E. Weiss, H. B. Rees, Jr., J. R. Hayes, *Nature* **213**, 1020 (1967).
8. E. Weiss, L. W. Newman, R. Grays, A. E. Green, *Infect. Immun.* **6**, 50 (1972).
9. The medium was developed by D. G. Evans in the laboratory of one of us (E.W.). It consisted of 35 ml of a semisolid phase and 15 ml of a fluid phase both containing 1.5 percent vitamin-free casamino acids, 0.12 percent yeast extract (Difco), 0.002 percent hemin chloride, and 1.49 percent TES buffer (Calbiochem) [*N*-tris(hydroxymethyl)methyl-2-aminoethane sulfonic acid], *p*H 7.8. The semisolid phase also contained 0.6 percent Noble agar. To the fluid phase were added 0.5 ml of ten times concentrated Eagle's MEM medium (minimum essential medium) and 0.15 ml of 5 percent bovine plasma albumin solution. The cultures were inoculated with 0.1 ml portions of heavy rickettsial suspensions, incubated at 37°C with flowing 5 percent $CO_2$ for 3 days, and the organisms were released from the agar with glass beads.
10. M. J. Taylor and C. B. Thorne, *J. Bacteriol.* **91**, 81 (1966).
11. F. J. Tyeryar, Jr., and W. D. Lawton, *ibid.* **100**, 1112 (1969).
12. J. Marmur, *J. Mol. Biol.* **3**, 208 (1961).
13. D. T. Kingsbury, *J. Bacteriol.* **94**, 870 (1967).
14. A. Kaplan, in *Fundamental Techniques in Virology*, K. Habel and N. P. Salzman, Eds. (Academic Press, New York, 1969), p. 487.
15. R. J. Britten, M. Pavich, J. Smith, *Carnegie Inst. Washington Yearb.* **68**, 400 (1970).
16. W. Szybalski, *Methods Enzymol.* **12**, 330 (1968).
17. K. Bott and B. Strauss, *Virology* **25**, 212 (1965).
18. D. B. Ritter and R. K. Gerloff, *J. Bacteriol.* **92**, 1838 (1966); F. J. Tyeryar, Jr., and W. D. Lawton, *ibid.* **104**, 1312 (1970).
19. G. R. Wyatt and S. S. Cohen, *Nature* **170**, 846 (1952).
20. W. H. Price, *Amer. J. Hyg.* **58**, 248 (1953).
21. S. B. Wolbach, J. L. Todd, F. W. Palfrey, *The Etiology and Pathology of Typhus* (Harvard Univ. Press, Cambridge, Mass., 1922); J. W. Vinson and H. S. Fuller, *Pathol. Microbiol.* **24** (suppl.), 152 (1961).
22. J. A. McKiel, J. Bell, D. B. Lackman, *Can. J. Microbiol*, **13**, 503 (1967); W. Burgdorfer and L. P. Brinton, *Infect. Immun.* **2**, 112 (1970); D. A. Wike, G. Tallent, M. G. Peacock, R. A. Ormsbee, *ibid.* **5**, 715 (1972).
23. Since this report was submitted S. Schramek [*Acta Virol. Prague* **16**, 447 (1972)] reported G + C contents, as determined by $T_m$, for *R. prowazeki*, *R. mooseri* (*R. typhi*), and *R. canada*. His values are in agreement with those presented here.
24. Supported in part by the Bureau of Medicine and Surgery, Department of the Navy, Research Task MR041.05.01-0007A6GI. We thank A. E. Green, L. W. Newman, and G. Tallent for technical assistance.

7 December 1972

## Applications of Artificial Intelligence: Relationships between Mass Spectra and Pharmacological Activity of Drugs

Abstract. *The possibility that the mass spectrum and pharmacological activity of a compound may be directly related has been explored with the help of various computer-based pattern-recognition techniques. The relationship appears to hold at least for tranquilizers and sedatives, and compounds with one or the other of these two pharmacological activities can thus be classified from their mass spectra with a high degree of accuracy.*

It has long been a basic assumption of organic mass spectrometry that the mode of fragmentation of an organic compound, which dictates all the details of its mass spectrum, is itself determined by the structure of the molecule (*1*). There is now much evidence in support of this view (*2*), which implies that the mass spectrum of a compound is in fact no more than a representation of its molecular structure, expressed in a complex and incompletely understood code.

Likewise, it has been known empirically for many years, and has always appeared to be entirely acceptable on an empirical basis, that the pharmacological activity of any molecule is dependent upon its structure, and that a change in the structure can lead to a change in the activity (*3*). A vast amount of effort has been expended upon studies of the interplay between these two properties of molecules (*4*).

In view of these two relationships, it appeared to us to be of some considerable interest to examine the one remaining possible correspondence, that be-

tween pharmacological activity and mass spectrum. If the latter is indeed a representation of the structure of the molecule, then comprehension of the way in which activity is related to it should follow from a careful study of only the mass spectra and the pharmacological activities. The structures of the molecules are not required in this exercise.

We now report on a series of experiments we have conducted with a view to testing this hypothesis. The mass spectral data used in this work were taken from the collection of mass spectra of drugs published by Sub-Committee VI of the American Society for Mass Spectrometry. The compounds whose spectra were used in this connection were mainly either sedatives or tranquilizers and are listed in Table 1. The initial objective of this study was to discover if a machine can learn to distinguish between sedatives and tranquilizers on the basis of their mass spectra alone. The following series of computer experiments have been conducted with a view to studying this question.

1) *The distance approach.* Mass spectra may be considered as points in $n$-dimensional space, $n$ being the number of mass units to be considered, whose coordinates are the ion intensities at each of the $n$ $m/e$ values ($m/e$ = mass to charge ratio). Given spectra $x = (x_1, x_2, \ldots x_n)$ and $y = (y_1, y_2, \ldots y_n)$, we can measure the similarity between them by calculating the Euclidean distance between them according to the following formula, in which $i$ refers to $m/e$ value and $I$ to intensity:

$$d_{x,y} = \left[ \sum_{i=1}^{n} (I_{i,x} - I_{i,y})^2 \right]^{1/2}$$

If $d_{x,y}$ is small, the implication is that the two spectra are similar to each other. Moreover, it is reasonable to assume that compounds which are in the same classification should have spectra that are close to each other, that is, that cluster together. Experiment 1 was designed to test the above assumption.

1a) *The K-nearest neighbor method.* The K-nearest neighbor method (*5*) simply involves the placing of an unknown spectrum into the same class as that of a majority of its K-nearest neighbors in $n$-dimensional space. When this method, with K = 1, was applied to the 66 compounds in Table 1, with each in turn used as an unknown and the remaining 65 as knowns, 55 compounds (83 percent) were classified correctly.

The technique of nonlinear mapping (*6*), a variation on the K-nearest neighbor method, involves the calculation of a two- or three-dimensional plot in which all the interpoint distances in $n$-dimensional space are preserved as well as possible. This cannot be done without introducing positional errors, and the algorithm used (*7*) works to minimize these errors. The human, viewing such a plot, makes judgments as to the classification of any point by a combination of K-nearest neighbor and distance criteria. Application of this approach to a data set containing the mass spectra of compounds 14 and 64 as unknowns and the remaining compounds in Table 1 as knowns resulted in only a moderate degree of clustering.

1b) *The weighting of peaks.* It has been pointed out (*8*) that data can be refined by weighting the variables. Mass spectra are readily susceptible to such treatment because some ions clearly contain more structural information than others. Given a spectrum ($x_1$, $x_2$, . . . , $x_n$), a weight $R_i$ can be associated with every $x_i$ transforming the spectrum to ($R_1x_1$, $R_2x_2$, . . . , $R_nx_n$). $R_i$ is the "Fisher ratio" (*8, 9*) of a particular ion of $m/e$ value $i$ and is calculated as follows:

$$R_i = \frac{(\bar{X}_{i1} - \bar{X}_{i2})^2}{\sigma_{i1}^2 + \sigma_{i2}^2}$$

where $\bar{X}_{i1}$ and $\bar{X}_{i2}$ are the average intensities of the ions of mass $i$ in class 1 and class 2, respectively, and $\sigma_{i1}$ and $\sigma_{i2}$ are the standard deviations of these intensities. The meaning of $R_i$ has been discussed in detail by Lee and Ting (*8*).

In this experiment, 64 drugs from Table 1 were chosen as knowns, and two, numbers 14 (P) and 64 (Q), were considered as unknowns. When Fisher ratios were used to weight the data, nonlinear mapping gave a plot in which two distinct clusters, the larger one containing both unknowns, were clearly discernible. This group of 51 compounds, which included P and Q, was reweighted, and the reduced set gave a new nonlinear mapping in which P and Q were in a cluster of 41 knowns. These 43 compounds were recycled once more to give the nonlinear mapping shown in Fig. 1, in which the two unknowns can be correctly classified by consideration of their K-nearest neighbors and the various interpoint

Table 1. Drugs employed in mass spectral artificial intelligence studies.

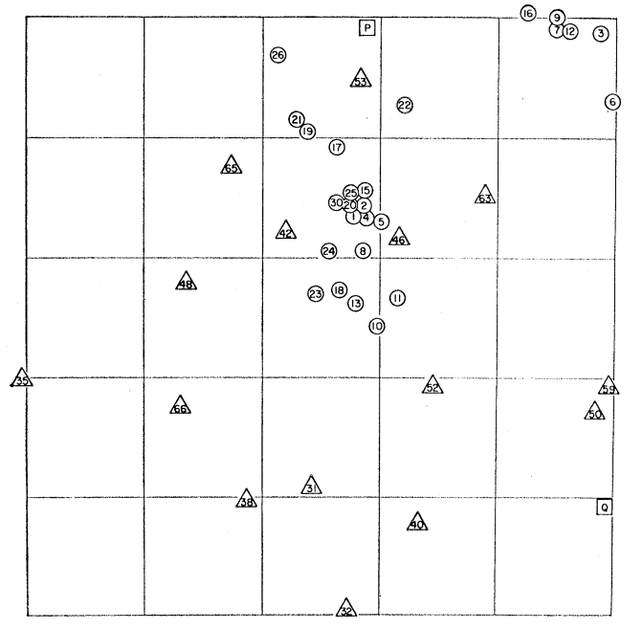
| Sedatives | Tranquilizers |
|---|---|
| 1. Mephobarbital | 31. Acetophenazine |
| 2. Phenobarbital | 32. Carphenazine |
| 3. Barbital | 33. Meprobamate |
| 4. Alphenal | 34. Mepazine |
| 5. Allobarbital | 35. Tybamate |
| 6. Butethal | 36. Methoxypromazine |
| 7. Amobarbital | 37. Trifluopromazine |
| 8. Aprobarbital | 38. Oxanamide |
| 9. Pentobarbital | 39. Acetylpromazine |
| 10. Secobarbital | 40. Pericyazine |
| 11. Butalbital | 41. Perphenazine |
| 12. Hexethal | 42. Mephenoxalone |
| 13. Talbutal | 43. Aminopromazine |
| 14. Cyclobarbital | 44. Dixyrazine |
| 15. Heptabarbital | 45. Methotrimeprazine |
| 16. Probarbital | 46. Ectylurea |
| 17. Hexobarbital | 47. Thioperazine |
| 18. Thiamylal | 48. Hydroxyzine |
| 19. Chloral betaine | 49. Prochlorperazine |
| 20. Ethchlorvynol | 50. Promethazine |
| 21. Ethinamate | 51. Chlorpromazine |
| 22. Phenaglycodol | 52. Methdilazide |
| 23. Acetylcarbromal | 53. Thioridazide |
| 24. Carbromal | 54. 3-Acetoxy chlorpromazine |
| 25. Glutethimide | 55. 8-Acetoxy chlorpromazine |
| 26. Methyprylon | 56. 3-Hydroxy-2-chlorpromazine |
| 27. Captodiamine | 57. 8-Hydroxy-2-chlorpromazine |
| 28. Methapyrilene | 58. Promazine |
| 29. Pyrilamine | 59. Propiomazine |
| 30. Tetrahydrocannabinol | 60. Triflupromazine |
| | 61. Thiopropazate |
| | 62. Chloprothixene |
| | 63. Deserpidine |
| | 64. Diazepam |
| | 65. Chlordiazepoxide |
| | 66. Buclizine |

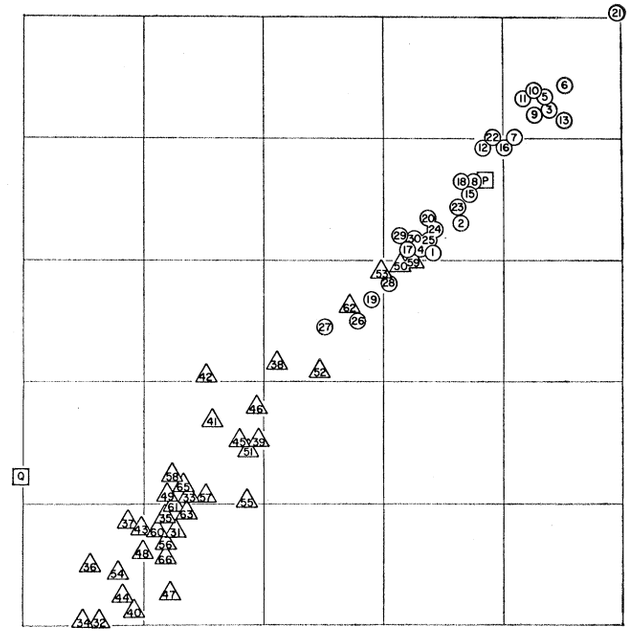
distances. In seven further experiments with this method, defining as unknowns compounds 1, 15, 18, 22, 27, 62, and 65 in turn, the correct classification was achieved in every case. The number of cycles necessary varied from one to four.

2) *The Fisher direction approach.* In experiment 1a, *n*-dimensional points were displayed by a nonlinear mapping technique. Another way of displaying *n*-dimensional points is to project them onto some two-dimensional plane. Sammon (*10*) has shown that given two classes of points, there can be extracted a plane, optimal in a certain sense, onto which these points can be projected. Since the determination of this plane requires the inversion of a large matrix (in this case a 606 × 606 matrix), the Fisher ratio concept was used to reduce the number of peaks involved in the computation. That is, the Fisher ratio of each peak was calculated and the 30 peaks with the highest Fisher ratios were selected. Experiment 2 used the Fisher direction method and was conducted in two parts.

2a) In this experiment, 64 compounds were used as knowns and compounds 14 (P) and 64 (Q) as unknowns. Based on the 64 compounds, the optimal plane was extracted according to Sammon's method (*10*), and all 66 compounds were projected onto this plane. It can now be observed from Fig. 2 that there are two clusters (sedatives and tranquilizers), and compounds 14 (P) and 64 (Q) can be classified unambiguously and correctly.

2b) This test was made with the 66 central nervous system (CNS) depressants in Table 1 as knowns and six hallucinogens as unknowns. These hallucinogens were lysergic acid diethylamide (*A*), mescaline (*B*), psilocin (*C*), psilocybin (*D*), dimethyl tryptamine (*E*), and 2,5-dimethoxy-4-methyl amphetamine (*F*). An optimal plane was first determined from the 66



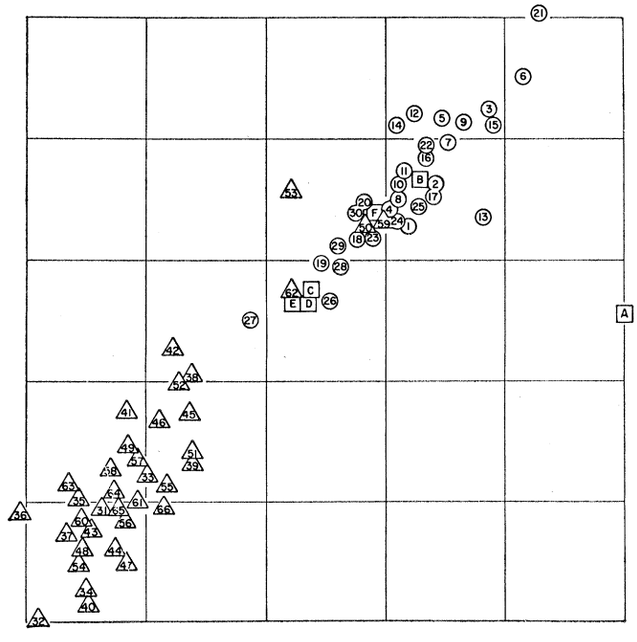




Fig. 1 (top left). Nonlinear mapping of weighted data; cycle 3.

Fig. 2 (top right). Fisher direction method applied to compounds in Table 1.

Fig. 3 (left). Fisher direction method applied to compounds in Table 1 with six hallucinogens added. *A*, Lysergic acid diethylamide; *B*, mescaline; *C*, psilocin; *D*, psilocybin; *E*, dimethyl tryptamine; *F*, 2,5-dimethoxy-4-methyl amphetamine.

knowns and then all 72 of the spectra were projected onto this plane. The interesting result of this test is shown in Fig. 3, in which it can be seen that four of the hallucinogens are found to be loosely distributed in the "gray area" between the two clusters and only two are graphed with the sedatives.

It is important to note that, after selection of the 64 compounds as knowns, as in experiment 2a, the correct categorization of the remaining compounds is by no means trivial, although occasionally some correct classifications can be made by using very simple rules. Half the sedatives are in fact barbituric acid derivatives whose spectra, structures, and activities are well known. That any of these can be correctly classified comes as no surprise, but that other, structurally diverse sedatives (compounds 20, 22, 24, and 25, for example) in Table 1 have been defined as unknowns and then correctly classified is noteworthy. Compounds 19 and 30 proved to be difficult to classify unambiguously. Both were classifiable as sedatives but with less certainty than in the other cases. It is of interest here to note that there is some uncertainty surrounding the definition of tetrahydrocannabinol (compound 30) as a sedative. Further, in view of the fact that the molecular weight of the sedatives is commonly below 250 while many of the tranquilizers have molecular weights above 250, a classification on this basis would seem simple. In practice, however, five of the tranquilizers have molecular weights below 250 and four of the sedatives have molecular weights above 250. More importantly, the 30 masses with the highest Fisher ratios include only one (low-ranking) member of mass greater than 250. Molecular weight is therefore of very little importance in the method developed in this way. The precise factors involved in the classification process are of great interest, and further work is needed to identify these.

Much further inquiry is suggested by this work. The pharmacological activity could, in principle, be identified by posing a series of consecutive, binary questions (for example, CNS active or not? If CNS active, depressant or stimulant? If depressant, sedative or tranquilizer? and so on) and attempting to answer these questions by using the methods described above. This would provide the basis of a general approach to the screening of compounds for pharmacological activity. A second, particularly interesting problem is that of deciphering the rules that the machine is developing in achieving the clustering. If these rules can be isolated and reexpressed in terms of mass spectral features, or even chemical structures, the way is clear to a relatively facile method of studying relationships between structure and pharmacological activity.

KAI-LI H. TING, R. C. T. LEE
G. W. A. MILNE, M. SHAPIRO
A. M. GUARINO
*Division of Computer Research and Technology, National Cancer Institute* and *National Heart and Lung Institute, National Institutes of Health, Bethesda, Maryland 20014*

### References and Notes

1. K. Biemann, *Mass Spectrometry, Applications to Organic Chemistry* (McGraw-Hill, New York, 1962).
2. H. Budzikiewicz, C. Djerassi, D. H. Williams, *Mass Spectrometry of Organic Compounds* (Holden-Day, San Francisco, 1967).
3. A. Burger, in *Medicinal Chemistry*, A. Burger, Ed. (Wiley-Interscience, New York, ed. 3, 1970), part 1, pp. 64–80; A. Cammarata and A. N. Martin, in *ibid.*, pp. 118–163.
4. Investment by the American pharmaceutical industry in this area has for some years been about 6 percent of their total R & D budget. [*The Drug Makers and Drug Distributors* (Government Printing Office, Washington, D.C., 1968)].
5. N. J. Nilsson, *Learning Machines* (McGraw-Hill, New York, 1965), pp. 119–120; B. R. Kowalski and C. F. Bender, *Anal. Chem.* **44**, 1405 (1972).
6. B. R. Kowalski and C. F. Bender, *J. Amer. Chem. Soc.* **94**, 5632 (1972); J. W. Sammon, *IEEE (Inst. Elec. Electron. Eng.) Trans. Comput.* **C18**, 401 (1969).
7. C. L. Chang and R. C. T. Lee, *IEEE (Inst. Elec. Electron. Eng.) Trans. Syst. Man Cybernet.*, in press.
8. R. C. T. Lee and K. L. H. Ting, in preparation.
9. R. A. Fisher, *Ann. Eugen.* **7**, 179 (1936).
10. J. W. Sammon, *IEEE (Inst Elec. Electron. Eng.) Trans. Comput.* **C19**, 826 (1970).
11. We are grateful to Drs. C. L. Chang and H. M. Fales for their advice and to Ms. V. A. Aandahl for her assistance in the collection and organization of the data used in this work.

16 November 1972; revised 5 February 1973 ■

## Visual Resolution and Experience: Acuity Deficits in Cats Following Early Selective Visual Deprivation

Abstract. *Cats reared during the first 5 months of life in environments that contain contours of a single orientation show a diminished ability to resolve gratings of the orthogonal orientation in later life. It is argued that these perceptual deficits result from changes in the organization of the visual cortex induced by the selected early visual input.*

Many of the properties of neurons in the visual cortex of cats and monkeys are influenced by the visual experience of the first 3 months of life (*1–4*). Dramatic demonstrations of this were provided by the studies of Hirsch and Spinelli (*2*) and Blakemore and Cooper (*3*), who restricted the visual input of each eye of young kittens to stripes of a single orientation (either vertical or horizontal). After several months of such rearing, all visual cortical neurons responded best to edges or bars having orientations similar to that of the contours to which each eye had been exposed. In the case of Blakemore and Cooper's kittens, which experienced contours of the same orientation in the two eyes during the first 5 months of life, no neurons at all could be found that had preferred orientations perpendicular to that of the contours in which they were reared. This distribution of preferred orientations is clearly very different from that found in normally reared cats, in which all orientations are equally represented (*5*).

It is possible that the properties of neurons in the human visual system are similarly susceptible to early visual experience. Humans with ocular astigmatism can be considered to have been "deprived" of sharp images of contours of certain orientations throughout the time that their refractive error was uncorrected. Even when the astigmatism is fully corrected with lenses, many adult astigmats show dramatically reduced acuity for gratings having the orientation that was habitually seen most blurred before correction (*6*). The argument that human visual resolution is influenced by early visual experience would be strengthened considerably if it could be shown that the physiological effects of selective visual deprivation in cats were accompanied by concordant deficits in visual acuity. In this report we provide evidence that cats selectively deprived of contours of certain orientations in early life do indeed show long-lasting and perhaps permanent deficits in their ability to perceive these contours as adults.

We reared three kittens using a procedure and environment virtually identical to that of Blakemore and Cooper (*3*); thereby we hoped to duplicate the physiological deficits they found. Each kitten was placed for 5 hours a day on a clear Plexiglas plate mounted in the middle of a cylinder 1.8 m high that had stripes of various widths but of a single orientation on the inside walls.