tip their heads 90° to one side (so that the physically vertical test stripes were retinally horizontal) and then answer the same five questions. All subjects had earlier been given practice filling out the checklists in response to four sample stimuli, three of which were actually tinted.

All eight subjects in the first experiment had previously seen the McCollough effect on a variety of test patterns. None of the 16 subjects in the second experiment had ever seen the effect or heard about its direction.

The main finding was that most subjects saw a McCollough aftereffect. On the last trial, 16 of the 24 subjects reported that they saw an appropriate aftereffect color on at least one part of the test figure, and six of the remaining eight named both appropriate colors when forced to "guess."

Thus, the McCollough effect occurs even when striped afterimages are precluded by flashing the colored grids randomly in two different locations. The phenomenon must therefore depend (as McCollough assumed) on neural units more complex than the individual retinal receptors that could (in principle) yield an ordinary afterimage.

Several possible objections to this conclusion are countered by various aspects of the procedure and results. For example, it is conceivable that striped afterimages could be produced by certain sorts of eye movements, by a preponderance of colored grids in one location, or by optical imperfections in the slides, the projectors, or the subjects' eyes. Most such objections would lead one to expect that the aftereffect would be at least as likely to be seen after a short adaptation period as after a long one, that it would wax and wane irregularly as adaptation time increased, and that subjects would often see the "wrong" colors on the test figure (whenever the afterimage overlapped the test pattern in an inappropriate way).

However, as total adaptation time increased, there was a steady growth in the strength of the aftereffect (Fig. 1). Only one subject ever did report, after seeing an aftereffect color on one trial, that the color looked weaker on any subsequent trial with head upright. And only one subject reported seeing an inappropriate color on the test pattern after being exposed to the colored grids.

Our findings show that afterimages cannot account for the McCollough effect. Must we then attribute it to color adaptation of edge detectors in the visual system? Although that inference is attractive, especially in view of the recent discovery of cells in the monkey's visual cortex that are differentially sensitive to both wavelength and orientation of a contour (*11*), the psychological data do not yet demand such a conclusion.

The perceptual phenomena observed to date could be ascribed to hypothetical sensory units much less highly structured than edge detectors are commonly thought to be—units that could not even report accurately on a contour's orientation, and thus could not individually yield a perception of an edge. The simplest such unit [Gibson and Harris (*5*) called it a "dipole"] would receive inputs from two nonconcentric areas of the retina. Given a population of fatiguable dipoles with some variation in spectral sensitivity and in spatial relation of receptive areas, very few additional assumptions are necessary to deal with all data on the McCollough effect (*12*). For example, if the dipoles responded to differences between light intensities on their two receptive areas, they would fire when a light-dark boundary fell between the two areas. This rudimentary model may help clarify which of the presumed properties of edge detectors and which physiological findings are actually relevant to the psychological data.

CHARLES S. HARRIS
ALAN R. GIBSON*
Bell Telephone Laboratories,
Murray Hill, New Jersey 07974

References and Notes

1. C. McCollough, *Science* **149**, 1115 (1965).
2. See, for example, D. H. Hubel and T. N. Wiesel, *J. Physiol.* (*London*) **148**, 574 (1959); J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, W. H. Pitts, *Proc. Inst. Radio Eng.* **47**, 1940 (1959); D. N. Spinelli, *Exp. Neurol.* **19**, 291 (1967); D. H. Hubel and T. N. Wiesel, *J. Physiol.* (*London*) **195**, 215 (1968).
3. Research on afterimages is reviewed by J. L. Brown in *Vision and Visual Perception*, C. H. Graham, Ed. (Wiley, New York, 1965), pp. 479–503.
4. A. Hajos, paper read at I. Zusammenkunft der Psychologen aus den Donauländern (1967); L. W. Teft and F. T. Clark, *Psychon. Sci.* **8**, 265 (1968); G. Murch, paper read at convention of the American Psychological Association (1968); L. Fidell, thesis, University of Michigan (1968); T. Sorenson, unpublished results; C. McCollough, unpublished results; C. F. Stromeyer, III and R. J. Mansfield, unpublished results.
5. A. Gibson and C. S. Harris, paper read at convention of the Eastern Psychological Association (1968).
6. Hajos's subjects reported seeing aftereffect bars during dark intervals between presentations of adapting grids (*4*).
7. The colors were produced by gelatin filters that give a strong McCollough effect: Edmund Scientific Co. No. 818 for the red, and No. 471 combined with "cyan" for the green.
8. M. G. Saslow, *J. Opt. Soc. Amer.* **57**, 1030 (1967).
9. A shaded incandescent light, permitting the subject to read the answer sheet, remained on during the entire session. The luminance of the answer sheet was less than 0.1 mlam and was negligible elsewhere.
10. The bias in responses that preceded exposure to the colored grids in the first experiment was eliminated in the second by counterbalancing conditions: half the subjects saw green vertical and red horizontal adapting grids, and half saw the reverse; for half, the center test diamond was vertically striped, and for half, horizontally.
11. D. H. Hubel and T. N. Wiesel, *J. Physiol.* (*London*) **195**, 215 (1968).
12. C. S. Harris and A. R. Gibson, paper read at convention of the Psychonomic Society (1968).
13. We thank S. Sternberg and J. Krauskopf for suggestions and R. A. Payne for aid in designing the randomizing circuit.
* Present address: Department of Psychology, University of California, Santa Barbara, California 93106.

16 August 1968; revised 7 October 1968

Potassium Feldspar in Weekeroo Station, Kodaikanal, and Colomera Iron Meteorites

In a recent paper (*1*), a reference was made to our "erroneously [reporting] a feldspar of composition $Ab_{64}An_9Or_{27}$ in Weekeroo Station as potassium feldspar" (*2*). No explanation or alternative nomenclature was offered by these authors; however, they apparently meant that, since the feldspar in question has an Ab : Or molecular ratio of 2:1, it should correctly be referred to as alkali feldspar.

The electron microprobe analysis of this feldspar, as given in our paper, is an average of 5 to 15 spot analyses on each of ten very small grains. The K_2O content ranges from 0.90 to 11.5 percent by weight with a corresponding reciprocal Na_2O content of 9.8 to 3.1 by percent weight; the average is 4.9 percent K_2O and 7.6 percent Na_2O. Potassium-rich areas were too small ($<10\ \mu$) to obtain a complete analysis without interference from the host material. X-ray diffraction studies indicate a possible antiperthitic intergrowth of sodic plagioclase and potassium feldspar. Unfortunately, the extremely small grain size and scarcity of material did not allow us to obtain precise x-ray data.

Potassium-rich areas are small in volume compared to the much larger sodic plagioclase host. The important point is that potassium feldspar is present in the Weekeroo Station meteorite. We chose to refer to the average of all analyses of these particular grains as potassium feldspar simply to distinguish between potassium feldspar and plagioclase and in keeping with the

main topic of the paper. We regret any confusion that was created by our use of the term "K-feldspar" instead of x-ray antiperthitic alkali feldspar.

T. E. BUNCH
Space Sciences Division, Ames Research Center, National Aeronautics and Space Administration, Moffett Field, California 94035
EDWARD OLSEN
Department of Geology, Field Museum of Natural History, Chicago, Illinois 60605

References

1. G. J. Wasserburg, H. G. Sanz, A. E. Bence, *Science* **161**, 684 (1968).
2. T. E. Bunch and E. Olsen, *ibid.* **160**, 1223 (1968).

25 October 1968

Chromosomal Effect and LSD: Samples of Four

The analysis by Sparkes, Melnyk, and Bozzetti (*1*) on the effect of LSD in vivo on human chromosomes creates a misimpression, primarily because they neglect the effect of their very small sample sizes. Closely associated with the problem of sample size is their neglect of the distinction between statistical significance and substantive significance. The distinction, which has been made for years (*2*), still is frequently misunderstood.

Sparkes, Melnyk, and Bozzetti worked with three groups of four people each: controls, users of LSD, and people medically treated with LSD. About 225 lymphocytes from each of the 12 persons were examined, and a variety of kinds of chromosomal damage was observed. Four scoring schemes were used; for brevity we repeat here in Table 1 the results for only one. Then the Wilcoxon-Mann-Whitney test was repeatedly applied, and no statistically significant results (at the usual levels) were obtained.

Our major comment is that, in comparing two samples of size four, the substantive, real difference must be very large to have reasonable power, that is, to have a reasonably large probability of detecting the real difference. Therefore, a finding of no statistically significant difference does not by any means preclude the existence of a material real difference.

Before considering the question of power, we first mention another way of looking at the consequences of small sample sizes: confidence intervals. Let us assume that the sampled populations differ essentially only by translation; for example, let us assume that for some unknown number Δ the underlying distribution of cell percentages for users is the same as that for controls after adding Δ to each control percentage. Then the Wilcoxon-Mann-Whitney test is readily applied (*3*) to obtain confidence intervals for Δ. The results, at the 94.3 percent level of confidence (*4*), in percentage units, are (i) users minus controls, −5.5 to 6.7; (ii) medically treated minus controls, −3.3 to 2.3; (iii) users minus medically treated, −4.8 to 7.8. The first of the above, for example, says that the observed difference between users and controls is not surprising (5.7 percent significance level) if one were testing null hypotheses that the real difference lies between −5.5 and 6.7. It seems to us that real differences of 6 or 7 in percentage units might be quite important; such real differences are consistent with the observed data.

Similar conclusions are reached from the viewpoint of power. For example, if breakage-gap scores had negative exponential distributions, and a significance level of .057 were used, the null hypothesis would be rejected only about 60 percent of the time, even if users had an average breakage rate *six* times that of controls. For a significance level of .029, a corresponding percentage is only achieved with an average rate for users nine times that for controls (*5*). If the parent populations are normally distributed with common variance σ^2, it is notable that, for a significance level of .029, the probability of rejecting the null hypothesis for a difference in means of 2.5 σ is .682 (*6*). It may be seen from Table 1 that σ is quite substantial.

There are other difficulties in reaching conclusions from this set of data, and we mention three of them. First, there is no reason to think that the three samples are either random or from the same population of humans. Some differences are immediate; for example, the medically treated subjects range in age from 28 to 45, while the users age range is 19 to 24. The control ages go from 21 to 50. This problem of basic noncomparability may be inherent in studies of this kind, and we do not take the view that valid conclusions in such circumstances are impossible. Nonetheless, an extra measure of caution is necessary. Second, we do not know whether the cells were observed blindly, that is, with the observers in ignorance of the source of the samples. Since determination of cell aberration doubtless has some subjective elements, a lack of blindness might introduce bias. Third, two laboratories analyzed separate samples of blood from each person. The differences between the results from the two laboratories (which used different techniques) would be illuminating, since they would give an idea of variability stemming from both the blood sampling and from the laboratory techniques. Unfortunately, the only information given is that there was no significant difference between results from the two laboratories.

Table 1. Percentages of cells with breaks or gaps (*1*). Samples are ordered within themselves.

Sample	Broken cells (%)			
Controls	3.3,	4.8,	6.4,	7.1*
Users	0.9,	2.6,	3.4,	11.5
Medically treated	3.1,	3.7,	5.7,	7.1*

* The unrounded 7.1 for controls is slightly less than the one for medically treated.

WILLIAM H. KRUSKAL
SHELBY HABERMAN
Department of Statistics, University of Chicago, Chicago, Illinois 60637

References and Notes

1. R. S. Sparkes, J. Melnyk, L. P. Bozzetti, *Science* **160**, 1343 (1968).
2. E. G. Boring, *Psychol. Bull.* **15**, 335 (1919); W. H. Kruskal, in *International Encyclopedia of the Social Sciences*, D. L. Sills, Ed. (Macmillan and Free Press, New York, 1968), vol. 14, pp. 238–250.
3. L. Moses, in *Statistical Inference*, H. M. Walker and J. Lev, Eds. (Holt, New York, 1953), chap. 18.
4. Because the test statistic has a discrete distribution, it is difficult to use a conventional level like 95 percent. We chose the readily available level closest to 95 percent.
5. R. A. Shoruch, *Technometrics* **9**, 666 (1967).
6. W. J. Dixon, *Ann. Math. Statist.* **24**, 611 (1954).

12 September 1968

Sparkes *et al.* (*1*) tested the null hypothesis that there is no difference in chromosomal aberrations between users of LSD and nonusers. Their data indicated no significant difference in aberrations among test groups, hence the hypothesis was accepted. The probability level chosen in their test of significance specified the risk they were willing to take of rejecting the hypothesis if it were true (type I error). But there is also a risk of accepting a false hypothesis (type II error). The chance of making a type II error can be determined only for a specified difference between means. For any specified difference, however, it is possible to determine how many replications would be