rate of the time marker by starting the paper when a response pip is made, and moving it until the next time pip occurs.

Experimental periods such as stimulus presentations can be recorded by displacing a pen for the duration of the period. However, unambiguous assignment of events to the appropriate time period would require that the paper be moved slightly at the onset and offset of the experimental period.

Conference on Use of On-Line Computers

The Third National Conference on the Use of On-Line Computers in Psychology has been scheduled for October 31, 1973, prior to the Psychonomic Society meetings November 1-3. The Conference will be held at St. Louis University, Donald Tepas presiding. Persons interested in submitting papers or suggesting symposia should contact Peter G. Polson, Computer Laboratory for Instruction in Psychological Research, University of Colorado, Boulder, Colorado 80302.

Computer Program Abstracts/Algorithms

Computer program abstracts/algorithms will now be published in this Journal. Individuals interested in submitting such information should forward an original plus two copies of the typed material to the Editor. With acceptable copy, publication lag should not exceed 8-12 weeks.

Notes

A decision rule for deleting descriptors in naturalistic observational studies of behavior

PATRICK R. HARRISON*

University of Tennessee, Knoxville, Tennessee 37916

Recent work in child ethology has emphasized exhaustive descriptive studies of behavior. The concepts used as stimulus sets for counting have explicit reference to movements, postures, gestures, and expressions. These descriptive studies have generated a new set of questions concerning the construction and analysis of descriptive data sets. One of these questions concerns the assessment of interrater reliability. Some descriptors occur with such low frequency that an independent assessment of interrater reliability cannot be made. It is also clear that as the number of descriptors used in a particular study increases it becomes less feasible to independently assess interrater reliability for each descriptor. It is suggested that the average interrater reliability coefficient can be used to generate rules for systematically dealing with these problems.

*Requests for reprints should be sent to Patrick R. Harrison, who is now at Hope College, Holland, Michigan 49423. The author gratefully acknowledges the comments and suggestions of J. Reynierse in the preparation of this paper.

An initial step in observational studies such as those done by the child ethologists (Jones, 1972; McGrew, 1972) is to construct a preliminary classification system consisting of a limited number of unordered categories of behavior, such as locomotion, each with a comprehensive set of descriptors, such as walk, run, and skip. A critical problem in refining this preliminary classification system is the assessment of interrater reliability.[1]

The probability of occurrence of some descriptors may be quite low, making impossible an independent assessment of interrater reliability. As McGrew stated, "Thirty patterns [descriptors] could not be assigned coefficients, in most cases because of infrequent occurrence. [An arbitrarily chosen minimum of seven recorded instances was required for computing a coefficient] [1972, p. 38]." It is also patently clear that as the number of descriptors used in an observational study increases, it becomes less and less feasible to calculate an interrater reliability coefficient for every descriptor.

This paper proposes a quantitative decision rule that relates the average interrater reliability for the descriptors under each major category to their proportion of occurrence such that raw frequency count alone is not used as a basis for the inclusion or exclusion of a descriptor.

The decision rule is based on Fisher's Z formula (Steele & Torrie, 1960, pp. 189, 452). Utilizing a version of his formula:

$$Z = (Z\rho_o - Z_r)/(1/\sqrt{n-3}),$$

and solving for n, we have:

$$n = \frac{Z^2 + 3}{(Z\rho_o - Z_r)^2}.$$

Constructing the rule requires converting the average interrater reliability coefficient (r) for the descriptors under each category, for a null hypothesis (ρ_o) and for a significant level (α) into Fisher Z scores.

Selecting a value for α is usually done by convention. Assume that we are interested in testing significance at the .05 level of confidence. We can then replace α by $Z = 1.96$. Rewriting our equation, we have:

$$n = (1.96)^2 + 3/(Z\rho_o - Z_r)^2,$$

$$= 6.842/(Z\rho_o - Z_r)^2.$$

Selection of ρ_o is often arbitrary in the sense that it is conventional to uncritically use $\rho_o = .00$. Such an approach fails to account for both the idiosyncrasies and consequences of a particular study. A more critical procedure for establishing ρ_o requires that we establish a baseline value for the coefficients we expect to find in a

Table 1
Minimum n Required for Each Descriptor for Inclusion in the Analysis

Category*	Average Interrater Reliability	Z_r**	n† ρ_o = .00	ρ_o = .25	ρ_o = .50
1	.63	0.741	13	29	185
2	.72	0.908	9	17	54
3	.83	1.188	5	8	17
4	.90	1.472	4	5	9
5	.93	1.658	3	4	6
6	.98	2.298	2	2	3

*Based on a hypothetical data set.
**$Z_r = 0.5 \ln (1 + \rho)/(1 - \rho)$.
†n was calculated for ρ_o = .00, .25, and .50. The final form of the equation for obtaining n was $n = 6.842/(Z\rho_o - Z_r)^2$. n was rounded up to the next whole number.

particular study. This can be based, presumably, on a number of explicit and implicit criteria. McGrew (1972, p. 24), for example, established .70 as a baseline value. Any reliability coefficient less than .70 was considered suspect. Setting the established baseline value equal to Z_r and given the overall N used in the reliability study, we can then calculate ρ_o using the above formula.

Table 1 contains the results of applying this decision rule to a set of hypothetical behavior categories for three possible values of ρ_o (ρ_o = .00, ρ_o = .25, ρ_o = .50). Thus, for Major Category 4 (for ρ_o = .50), a minimum of nine occurrences of any of its descriptors would be the required cutoff for inclusion in subsequent analyses in terms of the initial reliability estimate. If n is less than nine, we risk using a descriptor for which reliability could not be assessed in terms of the original average reliability estimate.

This rule allows us to reduce systematically a data set without necessarily excluding descriptors that have a low frequency count and without subjectively defining "how small is small." When calculating reliability, it provides a cutoff less subjective than an a priori statement such as "any descriptor with an n less than 7 was arbitrarily excluded." In terms of the problem of calculating interrater reliability coefficients for large descriptor sets, it suggests that a random sample of descriptors under each major category would be tested for reliability and that the results could be used for evaluating the remaining descriptors under each major category.

REFERENCES

Jones, N. B. (Ed.) *Ethological studies of child behavior.* Cambridge: University Press, 1972.
McGrew, W. C. *An ethological study of children's behavior.* New York: Academic Press, 1972.
Steele, R. G. D., & Torrie, J. H. *Principles and procedures of statistics.* New York: McGraw-Hill, 1960.

NOTE

1. Intrarater reliability and measurements of internal consistency were not considered in this paper.

(Received for publication November 7, 1972; revision received January 27, 1973.)

A nonsurgical method for restraining the head during behavioral tests in primates*

ISRAEL A. GOLDBERG, EILEEN KOWLER and EDWARD TAUB
Institute for Behavioral Research, 2429 Linden Lane Silver Spring, Maryland 20910

Behavioral and neurophysiological experiments with alert Ss sometimes necessitate the restriction of head movements. Biteplates, headrests, and chinrests are commonly employed with human Ss, but require cooperation and therefore are not suitable for use with animals. The present article considers a number of head-restraint techniques employed with monkeys, as well as some problems associated with these techniques, and then describes a nonsurgical method devised to effect restraint of head movements in an experiment on visual-motor coordination.

Surgical techniques have been described for restraining an awake monkey's head while electrical activity is recorded from the brain (Sheatz, 1961), while the animal is engaged in performing trained movements of the forelimbs in a prism-adaptation study (Bossom, 1964), and while single-unit activity is recorded from Ss engaged in performing trained movements of the wrists (Evarts, 1968a, b). Hobson (1972) described a similar method for restraining cats during single-unit recording. All of these techniques require drilling keyhole-shaped slots through the calvarium, passing stainless steel screws head down through the slots, attaching the screws to the skull with lock washers and nuts, and further securing the screws in the slots with dental cement. Metal plates are then fastened to the exposed shafts of the implanted screws and, when restriction of head movements is desired, the plates are secured with rigid couplings to appropriate fixed points (e.g., to the frame of the apparatus in which S is seated).

There are several problems associated with restraining the head by means of screws implanted in the skull: (1) Implementation of the technique requires the capability for surgery not available in many laboratories. (2) The surgery and anesthesia required place the animal at a risk which, though minor, may be more than is desired for animals made especially valuable by either long training or successful recovery from life-endangering procedures. (3) The possibility of infection renders long-range studies questionable.

A number of techniques were examined before arriving at an acceptable alternative to surgical restraint of head movement in rhesus monkeys. First, a variety of rigid, tight-fitting helmets, to which screws could be affixed, were developed. It was felt that in this manner the problems of securing screws directly into the skull would be circumvented. Individual helmets were made out of thermoplastic neoprene (Johnson & Johnson).

*This research was supported by NIMH Grant IR01MH18821. The authors acknowledge the many helpful suggestions and assistance of Gilbert Barro, Lorraine Biros, and Philip Perrella.