Pharmacology Biochemistry & Behavior, Vol. 10, pp. 161–164. Printed in the U.S.A.

Videotaping: The Evaluation of Stereotypic Effects of Antiparkinsonian Agents[1,2]

RICHARD E. WILCOX, WILLIAM H. RIFFEE[3] AND ROBERT V. SMITH

Department of Pharmacology and Drug Dynamics Institute, College of Pharmacy
The University of Texas at Austin, Austin, TX 78712

(Received 29 August 1978)

WILCOX, R. E., W. H. RIFFEE AND R. V. SMITH. *Videotaping: The evaluation of stereotypic effects of antiparkinsonian agents.* PHARMAC. BIOCHEM. BEHAV. 10(1) 161–164, 1979.—Two experiments designed to assess the utility of a videotaping methodology in laboratory investigations of the stereotypic effects of putative antiparkinsonian agents are presented. The technique minimizes experimental bias through the use of "blind" ratings and increases the rating accuracy through the use of within- and between-animals controls. Permanent videotape records allow training of raters to high efficiency, and direct comparisons of different rating scales under identical experimental conditions.

Videotape Stereotypy Antiparkinsonian agents Apomorphine N-*n*-propylnorapomorphine Cage climbing

RATING SCALES are often the most practical means for assessing drug-induced responses in laboratory and clinical settings. For example, antiparkinsonian agents such as apomorphine produce a characteristic stereotypic response in rodents (Fig. 1) that is easily differentiated from other ongoing behavior and which can be evaluated through the use of ratings [1, 3, 4]. Especially when several responses are rated simultaneously in a number of subjects, the limited time for visual scanning imposes a burden upon the observers [7, 11, 13]. In addition, where "blind" ratings are not reported it may be especially difficult for complete objectivity to be maintained, since the observers presumably possess a knowledge of the expected drug-induced responses. Alternatively, with the commonly used methodologies it can be difficult to set up a protocol to insure accuracy and objectivity of ratings, since several observers would have to be present for each set of ratings [11]. Others have made effective use of videotaped interviews coupled with double-blind testing/rating conditions in the evaluation of drug-induced modifications of tardive dyskinesia [5]. Yet this technique appears not to be widely used to assess activity and stereotypic behavior in the laboratory setting. Stereotypic responding is widely used as a model system for the evaluation of agents effective in disorders of the sensorimotor system such as parkinsonism and tardive dyskinesia [9]. Stereotypy is also a potent behavioral tool for evaluating specific dopaminergic activity as a correlate of biochemical models of dopamine receptor function [10,12]. Therefore, a technique which improved the accuracy of ratings of stereotypy would be of general applicability [7, 13-15].

We now report two experiments which have evaluated the relative accuracy of "classical" and "videotape" methods of rating stereotypic responding. Included in this evaluation are the cost-benefit factors associated with each procedure as revealed by the answers to several questions. First, how convenient is each method to apply in a particular experimental situation? Do the relative advantages/disadvantages of the two methods change with increasing complexity of the rating task? Second, how difficult is it to compare drug-induced behavior with behavior occurring after the injection of control solutions using each methodology? Third, what is the interrater reliability with each of the two procedures and how convenient is it to determine this reliability routinely as a standard part of the experimental protocol? Experiment 1 compared the interrater reliability of videotaped ratings of stereotypic responding induced by two putative antiparkinsonian agents for two rating scales. Experiment 2 compared the interrater reliability of ratings of stereotype under "videotape" and "classical" conditions of observation.

METHOD

Animals

Ninety-five experimentally naive male albino mice (20–30 g; Charles River) were housed in group cages until just prior to behavioral testing. Food and water were available ad lib except during testing. A 12 hr light-dark cycle was main-

[1]This work was supported by Grant NS-12259 (RVS) and research fellowship NS-06114 (REW), National Institute of Neurological and Communicative Disorders and Stroke. The authors are grateful to the Sterling-Winthrop Research Institute (Rensselaer, NY) for a generous sample of N-*n*-propylnorapomorphine.

[2]The authors acknowledge the excellent technical assistance of Ms. Julie Anderson.

[3]Requests for reprints should be mailed to: Dr. Riffee, Department of Pharmacology, College of Pharmacy, The University of Texas at Austin, Austin TX 78712.

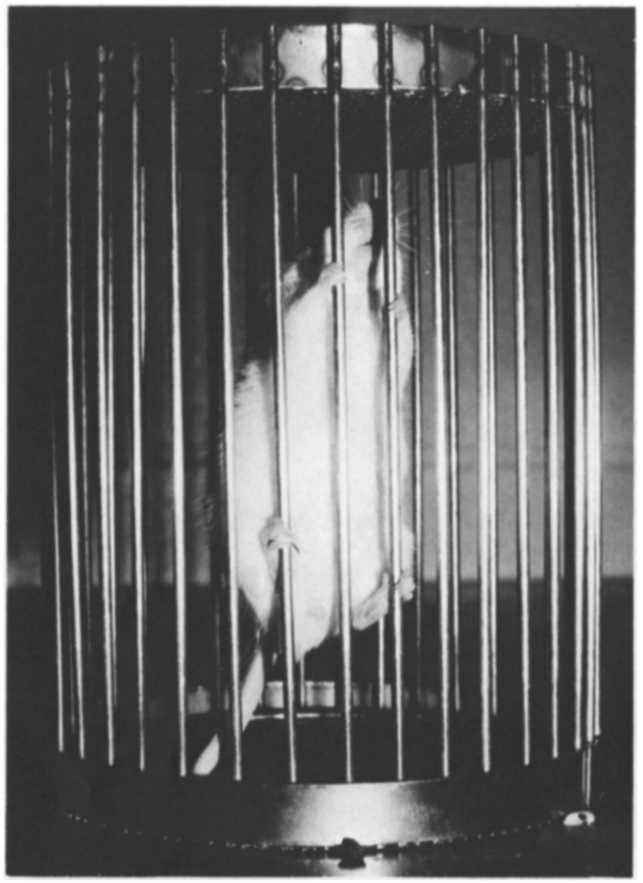

FIG. 1. Stereotypic activity induced by putative antiparkinsonian agents. Animal shown is exhibiting a maximal cage climbing ("verticalization") response to apomorphine. See text.

tained throughout the experiments (lights on 7 a.m.–7 p.m.). All testing was carried out between 8 a.m. and 6 p.m. Drugs administered to the mice included apomorphine hydrochloride hemihydrate (APO; 0.5–10 mg/kg) and N-*n*-propylnorapomorphine hydrochloride (NPA; 0.5–10 mg/kg), both of which induce dose-related (Riffee *et al.* [9,10]) cage climbing *via* dopamine receptor stimulation (Wilcox *et al.* [16], Wilcox *et al.*, submitted for publication). An initial interest in the investigation was possible potency differences between APO and NPA. Isotonic saline was used for control animals. All drugs were freshly prepared and administered intraperitoneally.

Apparatus

A black and white video camera (Sony), Trinitron television (Sony), and betamax videocassette recorder (Sony SL0320) were used to monitor, store and rate the cage climbing behavior. An automatic videotaping capability is possible by connecting the videocassette recorder to an automatic timer (Dayton) preset for the desired intervals.

Animals were placed in cylindrical cages, 12 cm dia., 14 cm high, constructed with vertical bars 2 mm dia., 1 cm apart, surmounted by fine wire mesh [6,8] for a 2 hr period of habituation to the environment to minimize exploratory behavior. Immediately prior to drug administration, the behavior of the animals was recorded on videotape (two 15 sec camera scans per animal). The animals were injected with one of the solutions indicated above and videotaping of post-drug behavior commenced exactly 15 and 30 min after injection for each animal with postdrug videotaping carried out exactly as the preinjection scans. Each videotape was coded by date, tape number, and the terms "preinjection," "15 min," and "30 min" while the identity of the solutions used for injection was recorded separately for decoding after the completion of the ratings.

Rating Scales

Stereotypic behavior was evaluated by means of two rating scales. Rating scale I [8]: 0=four paws on cage floor (behavior indistinguishable from that of normal, or placebo treated animal); 1=two paws holding the cage wall (minimal stereotypic response); 2=four paws holding the wall (maximal stereotypic response). Rating scale II (2): 0=behavior indistinguishable from that of normal or placebo treated animals: 1=increased locomotor activity, pacing, tail stiffening, some sniffing; 2=addition of occasional intermittent clinging to the sides of the cage with forepaws; 3=addition of intermittent clinging with hindpaws as well as forepaws; 4=intense, virtually uninterrupted clinging to the sides or top of the cage with all paws.

In Experiment 1, two trained observers independently scored the videotape responses of 73 mice using rating scale I. One week later the videotapes were again rated using scale II. Ratings of the behavior were made without knowledge of the drugs given ("blind") by using coded videotape. The behavior of the experimental animals after drug injection was compared to initial pre-drug behavior (within-animals control), and to the behavior of saline-injected animals (between-animals control).

RESULTS

Table 1 presents the results of the first experiment in which ratings are compared for cage climbing scores assigned by the two observers. The value of γ=for the comparison of the ratings assigned by the two observers using rating scales I and II is significant at the $p<0.001$ level. The Kendall coefficient of concordance, ω [17] was computed for the ratings of the 146 responses on each of four variables—V1=ratings of observer one using scale I; V2=ratings of observer one using scale II; V3=ratings of observer two using scale I; V4=ratings of observer two using scale II—and was found to be significant at $p<0.001$. These results demonstrate that the videotape methodology as presented here allows high reliability to be achieved in assigning stereotypy scores. Table 1 also presents the correlations between cage climbing scores assigned using rating scale I vs. the scores assigned using rating scale II. Their values (0.81 and 0.88 for observer one and two, respectively) suggest the consistency in ratings obtainable.

In the second experiment the cage climbing induced by four doses of apomorphine (1.25, 2.5, 5.0 and 10 mg/kg; n=6 per group) was assessed under two different conditions. Under condition one, stereotypy scores were assigned independently by directly observing the animals at 15 and 30 min after injection by one team of two observers who were aware of the doses of apomorphine used. The behavior of the animals was also recorded on videotape at this time for use in the second portion of the experiment. Under condition two, the videotaped behavior was rated three times at intervals of

TABLE 1

COMPARISON OF SCORES ASSIGNED BY TWO INDEPENDENT OBSERVERS TO STEREOTYPED BEHAVIOR RECORDED ON VIDEOTAPE. RATINGS OF CAGE-CLIMBING BEHAVIOR WERE ASSIGNED WITHOUT KNOWLEDGE OF DRUGS GIVEN USING CODED VIDEOTAPES ("BLIND") AND TWO RATING SCALES DESCRIBED IN THE TEXT. N=146

Comparison	Kendall Rank Correlation Coefficient (γ, tau)	Spearman Rank Correlation Coefficient (r_s)	Pearson Correlation Coefficient (r)
(1) Scale I: Ratings of Observer no. 1 vs those of Observer no. 2	.87*	.92*	.94*
(2) Scale II: Ratings of Observer no. 1 vs those of Observer no. 2	.87*	.92*	.93*
(3) Observer no. 1 Ratings using Scale I vs Scale II	.81*	.89*	.89*
(4) Observer no. 2 Ratings using Scale I vs Scale II	.88*	.94*	.92*

*Significant $p<0.001$.

TABLE 2

COMPARISON OF CAGE-CLIMBING SCORES ASSIGNED UNDER CONDITIONS OF DIRECT OBSERVATION OF THE BEHAVIOR AND OF RATING THE VIDEOTAPED BEHAVIOR. RATINGS OF STEREOTYPIC BEHAVIOR WERE ASSIGNED UNDER TWO CONDITIONS USING A RATING SCALE DESCRIBED IN THE TEXT: DIRECT OBSERVATION AND VIDEOTAPED RESPONSES. RANK CORRELATION COEFFICIENTS, SPEARMAN (r_s) AND KENDALL (γ) FOR STEREOTYPY SCORES ASSIGNED BY PAIRS OF OBSERVERS TO CAGE-CLIMBING BEHAVIOR FOLLOWING APOMORPHINE INJECTIONS (1.25, 2.5, 5.0, AND 10 MG/KG) TO 24 MICE ARE SHOWN. EACH SCORE REPRESENTS THE AVERAGE CAGE-CLIMBING SCORE ASSIGNED AT 15 AND 30 MIN AFTER DRUG INJECTION. N- 24

	Kendall Rank Correlation Coefficient (γ, tau)	Spearman Rank Correlation Coefficient (r_s)
Condition I. Direct observation of behavior	.85*	.93*
Condition II. Videotape ratings	.97*	.99*

Significant $p<0.001$.

approximately one week (as in Experiment 1) by a different team of observers.

Table 2 presents the results of the second experiment. The lowest amount of agreement between the ratings assigned by the observers was found for the condition of direct observation (condition I; $\gamma=0.85$). Three independent ratings of the coded videotapes led to virtually perfect agreement between ratings assigned by the observers (condition II; $\gamma=0.97$). Assuming that a high degree of consistency between ratings of behavior assigned independently by two observers is a reflection of rating accuracy, these results demonstrate that the use of the videotape methodology allows high accuracy for stereotypy scores. We performed an analysis of variance for ordinal data (Kruskal-Wallis) upon the cage climbing scores assigned under conditions of direct observation (condition I) and after the third videotape rating (condition II). The chi-square for condition I was 9.0187 with a significance of $p<0.03$. When the same analysis was performed on the scores assigned under condition II the chi-square was 11.0853 with a signifiance of $p<0.01$. These results suggest that the videotape methodology, by reducing the variability of the data, is more likely to tease out true differences than is a methodology based upon direct observation of the behavior.

DISCUSSION

Complex rating scales are common in many current investigations of stereotypy. When using videotape equipment connected to an automatic timer it takes less time to videotape behavior than to rate it directly. While this does not include the time spent in rating, the actual rating requires only about 15 min per group of animals in order to achieve an interrater reliability of $\gamma=0.85$ or better. The rating time increases with the complexity of the rating task. For example, drugs such as apomorphine and N-*n*-propylnorapomorphine induced a complex of behavioral responses in mice and rats including gnawing, head bobbing, sniffing, and locomotor activity changes as well as cage climbing (Wilcox *et al.*, in preparation). Accuracy of ratings may still be increased to virtually any level designated by the researcher merely by rating the tapes until the criterion is reached since 30 sec of videotape can be translated into as many minutes of rating time as desired. Since a permanent record of the behavior is maintained, interrater reliability can be determined without the difficulty of scheduling experiments with several observers present, as a routine part of the experimental protocol, which we recommend. Since it is axiomatic that the conclusions based upon a statistical analysis can be no more valid than the initial observations, the value of a methodology which insures maximal accuracy and minimal bias in assigning behavioral scores becomes apparent. The videotape method discussed here provides an opportunity for accurate ratings of the behaviors of interest, for allowing extreme flexibility in inter-behavioral and inter-drug comparisons, for eliminating possible sources of experimenter bias and for the assessment of new rating procedures under identical experimental conditions. All these factors will increase the value of stereotypy as a model for evaluating putative antiparkinsonian agents.

REFERENCES

1. Anden, N.-E., A Rubenson, K. Fuxe and T. Hokfelt. Evidence for dopamine receptor stimulation by apomorphine. *J. Pharm. Pharmac.* **19:** 627–629, 1967.
2. Baldessarini, R. J., N. S. Kula, K. G. Walton and R. J. Borgman. Behavioral effects of apomorphine and diisobutyrl-apomorphine in the mouse. *Psychopharmacology* **53:** 45–53, 1977.
3. Colpaert, F. C., W. F. M. Van Bever and J. E. M. F. Leysen. Apomorphine: Chemistry, pharmacology, biochemistry. *Int. Rev. Neurobiol.* **19:** 225–268, 1976.
4. Ernst, A. M. Mode of action of apomorphine and dextroamphetamine on gnawing compulsion in rats. *Psychopharmacologia* **10:** 316–323, 1967.
5. Gerlach, J., T. Rye and P. Kristjansen. Effect of baclofen on tardive dyskinesia. *Psychopharmacology* **56:** 145–151, 1978.
6. Marcais, H., P. Protais, J. Costentin and J. Schwartz. A gradual score to evaluate the climbing behavior elicited by apomorphine in mice. *Psychopharmacology* **56:** 233–234, 1978.
7. Peachley, J., B. Rogers, J. Brien, A. Maclean and D. Rogers. Measurement of acute and chronic behavioral effects of methamphetamine in the mouse. *Psychopharmacology* **48:** 271–275, 1976.
8. Protais, P., J. Costentin and J. Schwartz. Climbing behavior induced by apomorphine in mice: a simple test for the study of dopamine receptors in striatum. *Psychopharmacology* **50:** 1–6, 1976.
9. Riffee, W. H., R. Wilcox and R. V. Smith. Stereotypic and hypothermic effects of apomorphine and N-*n*-propylnorapomorphine. *Eur. J. Pharmac.*, in press, 1978.
10. Riffee, W. H., R. Wilcox and R. V. Smith. Parkinsonism and apomorphine: Stereotypic and hypothermic effects of apomorphine and N-*n*-propylnorapomorphine in mice—antagonism by neurotransmitter receptor blockers. *Pharmacologist* **20:** 223, 1978. Abstracts of the Joint Meeting, American Society of Pharmacology and Experimental Therapeutics, and Society of Toxicology, Houston, Texas, August 1978, p. 223.
11. Schoenfeld, D. I., J. L. Neumeyer, W. Dafeldecker and S. Roffler-Tarlov. Comparison of structural and stereoisomers of apomorphine on stereotyped sniffing behavior of the rat. *Eur. J. Pharmac.* **30:** 63–68, 1975.
12. Snyder, S. H., I. Creese and D. R. Burt. The brain's dopamine receptor: labelling with [^{3}H]-dopamine and [^{3}H]-haloperidol. *Psychopharmac. Commun.* **1:** 663, 1975.
13. Schreiber, H., R. Bell, L. Conely, M. Kufner, J. Palet and L. Wright. Diminished reaction to a novel stimulus during amphetamine withdrawal in rats. *Pharmac. Biochem. Behav.* **5:** 687, 1976.
14. Quock, R. and A. Horita. Differentiation of neuropharmacological actions of apomorphine and d-amphetamine. *Pharmac. Biochem. Behav.* **5:** 627–631, 1976.
15. Vandlergugten, J. and J. Slangen. Release of endogenous catecholamines from rat hypothalamus *in vivo* related to feeding and other behaviors. *Pharmac. Biochem. Behav.* **7:** 211–219, 1977.
16. Wilcox, R. E., W. H. Riffee and R. V. Smith. Pharmacological effects of apomorphine and N-*n*-propylnorapomorphine in the mouse. *Fedn Proc.* **37:** 855, 1978. Abstracts of the 62nd Annual Meeting, Federation of American Societies for Experimental Biology, Atlantic City, NJ, April, 1978.
17. Winer, B. J. *Statistical Principles in Experimental Design*. New York: McGraw-Hill, 1971.