

Chapter IV

EXPERIMENTAL DESIGN

The purpose of the experiment in which psychedelic drugs were administered in a religious context was to gather empirical data about the state of consciousness experienced. These data form the basis for a comparison with the typology of the mystical state of consciousness which has been presented above. Tape recordings, written accounts, questionnaires, and personal interviews were used to collect the data.

Non-drug Factors

Although descriptions in the psychopharmacological literature of the effects of these substances (LSD, mescaline and psilocybin) vary from "model psychosis" to "transcendental experience", investigators seem to agree that an alteration in the usual state of consciousness is produced.

Researchers who report "transcendental" experiences in their subjects claim that set and setting are important factors.¹

¹For a general statement of this position see T. Leary's article, "How to Change Behavior," in Clinical Psychology, ed. G.S. Nielsen (Proceedings of the XIV International Congress of Applied Psychology, Vol. IV; Copenhagen: Munksgaard, 1962), pp.62-64. A more detailed account of the application and implications of this approach from the work of Leary's group is given by R. Metzner, G. Litwin, and G. Weil in The Relation of Expectation and Setting to Experience with Psilocybin: A Questionnaire Study (Dittoed by Harvard

Set is defined as the personal expectation, mood, mental attitude and past experience of the subject; setting is the external environment and atmosphere and includes the expectations of the investigator. Trust and confidence in both the administrator and the situation are emphasized as crucial. In this view, the drug is seen as the necessary means by which this kind of potential experience may be actualized in a person who is properly handled and prepared.

The kinds of experiences reported by those who have studied the use of these substances in religious ceremonies seem to support this theory.² Both the setting and psychological expectation are conducive to an experience of great positive significance for the participant. The participant knows what the procedure of the ceremony will be, and feels

University Department of Social Relations) pp. 1-30. The method pioneered by A.M. Hubbard and used by the Canadian investigators, N. Chwelos, D.B. Blewett, C.M. Smith, A. Hoffer, H. Osmond, J.R. MacLean, D.C. MacDonald and U.P. Byrne, is explained in detail in Chwelos and Blewett's Handbook for the Therapeutic Use of Lysergic Acid Diethylamide 25 (to be published), pp. 15-48. Sherwood, Stolaroff, and Harman have come to similar conclusions (op. cit. p. 69).

²As was mentioned in Chapter II, the following have been participant-observers in Indian ceremonies in which sacred mushrooms or peyote were eaten: Wasson and Wasson (op. cit. pp. 287-316), Slotkin (Tomorrow Magazine, Vol. IV, No. 3), pp. 64-70, and Osmond (Tomorrow Magazine, Vol. IX, No. 2, pp. 105-125).

at ease as part of a group which undergoes the experience together.

The effects of set and setting were planned to maximize the possibility that mystical phenomena would occur. The assumption was made that for experiences most useful for comparison with the typology of mysticism, the atmosphere should be broadly comparable to that achieved by tribes who actually use natural psychedelic substances in religious ceremonies. The particular content and procedure of the ceremony had to be made applicable (i.e., familiar and meaningful) for the participants. Attitude toward the experience, both before and during, was taken into serious consideration in the experimental design. Preparation was meant to maximize positive expectation, trust, confidence, and reduction of fear. Setting was planned to utilize this preparation through group support and rapport, friendship, an open and trusting atmosphere, and previous acquaintance with the procedure of the experiment in order to eliminate, if possible, feelings of manipulation which might arise. The physical environment was a private chapel. There, on Good Friday, a two-and-one-half-hour religious service which consisted of organ music, four solos, readings, prayers, and personal meditation was attended by twenty Christian theological

students, some of whom had taken psilocybin prior to the service.

Choice of Drug

Psilocybin was chosen because its duration of action is only four to five hours, compared to eight to ten for LSD and ten to twelve for mescaline, in doses equivalent in potency. The practical problem of the time needed to supervise experimental subjects dictated that the use of psilocybin was more feasible. Claims of various differences in the effects of these three drugs have not been conclusively demonstrated, e.g., greater color and visual imagery with mescaline, or more unpleasant reactions with LSD.³ Any apparent advantage of psilocybin in terms of less preoccupation with imagery or generally more pleasant reaction is perhaps because psilocybin is newer and has not yet been researched as thoroughly. Also, because of the greater potency per

³Unger summarizes the evidence (op. cit., pp. 2-3 of his manuscript copy). The similarity between LSD and mescaline is stated by P. Hoch, H. Pennes, J. Cattell, Chemical Concepts of Psychosis, ed. M. Rinkel (New York: McDowell, 1958), p. 143. H. Isbell found similar effects produced by LSD and psilocybin ("Comparison of the reactions induced by psilocybin and LSD-25 in man," Psychopharmacologia, Vol. I 1959, p. 37). He also reported cross-tolerance which indicates the probability of a common pathway for LSD and psilocybin (H. Isbell et. al., "Cross-tolerance between LSD and psilocybin," Psychopharmacologia, Vol. II 1961, pp. 147-159).

unit weight of LSD, comparisons are not always between dosage effects of equivalent potencies. Higher doses of the same drug usually produce more intense effects, whether positive or negative. The oral dosage of psilocybin used was 30 mg. This dosage corresponds roughly to 150-200 micrograms of LSD or 500-750 mg. of mescaline. The controls received 200 mg. of nicotinic acid in identical capsules. This control substance produces transient vasodilation of blood vessels in the skin, especially of the face, and general relaxation.⁴ This was used to potentiate suggestion in the control subjects, all of whom knew that psilocybin produced various somatic effects, but none of whom had ever had psilocybin or any related substance before the experiment.

Recruitment and Pre-testing of Subjects

Subjects were student volunteers from a local theological seminary. They were recruited through a lecture on rehabilitation-experiments in which psilocybin was given to convicts at the Concord State Prison. Those students who wanted a personal experience with psilocybin met with the experimenter. At this meeting it was explained that psilocybin would be administered during a private Good Friday wor-

⁴L.S. Goodman and A. Gilman, The Pharmacological Basis of Therapeutics (New York: Macmillan, 1955), p. 1701.

ship service, and questions were answered about possible harmful effects and risks involved in taking an experimental drug. An effort was made not to overemphasize negative effects such as fear and psychotic-like symptoms which might produce a negative set. The California Psychological Inventory (CPI) as well as a medical-history form⁵ and a pre-drug questionnaire⁶ was completed by all volunteers. The medical-history form was similar to that used by the Harvard University Health Services for screening volunteers for the Harvard Psilocybin Project (sponsored by the Center for Research in Personality).

The pre-drug questionnaire was designed by the experimenter to assess personal religious background, training, and experience in a predominantly open-ended way. Questions included church affiliation, theological position, conversion experience, mystical or other religious experience, and devotional life. During the following week, each volunteer participated in a two-hour interview with the experimenter. During this time the medical history was reviewed

⁵See appendix A for a sample of the medical-history form.

⁶See appendix B for a sample of the pre-drug questionnaire.

with further questioning on any positive item which had been mentioned. Particular note was taken of metabolic diseases; previous consumption of drugs, alcohol, and tobacco; seizures; family and personal history of mental illness and psychotherapy; and symptoms of hysteria. Each person was asked to interpret the meaning of the proverb, "A rolling stone gathers no moss," in order to check on fundamental abstract reasoning power. A physical exam was given to evaluate general physical condition and to provide a baseline in case any physical complaints developed after the experiment. Suggestibility was tested by having the person stand with his eyes closed and imagine that a strong wind was pushing him backwards. This test was also a measure of the degree of resistance or cooperation with the experimenter.

Conditions which would have been grounds for rejection of a volunteer were cardiac impairment sufficient to limit exercise or contraindicate emotional strain, liver enlargement or history of liver disease (e.g., hepatitis or jaundice), history of psychosis, present psychological imbalance (especially depression), or current intensive psychotherapy or analysis for severe psychological problems. No volunteer needed to be rejected on these grounds.

After the physical exam and review of the medical

history, the pre-drug questionnaire was used as a guide in discussing the person's religious background and, especially, experience. Most of this time was spent in eliciting a description of the most meaningful religious experiences in the person's life. Care was taken not to suggest the basic phenomenology of the mystical experience to the subject, but rather to let him talk about his own experiences which he had mentioned in his questionnaire.

Preparation and Grouping of Subjects

Each person was given a written summary of the planned procedure of events on Good Friday and of the cooperation which was expected from him in collecting data during the days following. These points were discussed and questions answered. Suggestions were made for preparation by self-examination in depth, meditation, private devotional life, or reading of literature deeply meaningful to the person. These were suggestions only, and each person was told to prepare in the way which suited him best, but every effort was made to have him prepare, in as serious a manner as possible, for a meaningful experience.

When all the interviews had been completed, the data on each person were reviewed. Notes had been taken during each interview, and afterwards a general impression of the

person had been written by the experimenter.

Related data from the questionnaire supplemented by the interview were combined into categories which were scored with a 0-5 rating scale made relevant to each category. The categories fell into three main divisions: 1) Past religious experience, 2) Religious background and training, 3) General psychological makeup. Ratings on the categories in these groups, the CPI scores, and the interviewer's general impression of each person were used to match the twenty volunteers into ten pairs. (For a more detailed description of the data used in matching, and how the 0-5 scale was adapted to each category, see the next chapter.)

The twenty subjects were also divided into five groups of four persons each, on the basis of friendship and compatibility. Most of the subjects were in the same classes, ate together, and lived in the two dormitories which were next to each other on the campus. Groups were constituted without reference to matched pairs. There were only two different groups in which a pair of subjects were together. Each group was assigned two leaders who met with the group before the experiment.

These leaders were familiar with the effects and potentialities of psilocybin through personal experience and

observation of others in group sessions of the Harvard Psilocybin Project. Each of the five leaders with the most experience worked in a group with another leader who had not had quite as much experience in helping people through the experience. The two leaders who worked together knew each other and felt compatible working as a team. They knew from past experience the positive and negative possibilities of the drug experience, and their very presence was a reassuring factor to the subjects. The chief purpose of these leaders was to aid in creating a friendly and trust-filled set and setting which, it was hoped, would maximize the potential for positive experience, and to manage with confidence any disturbing reactions which might occur. There was one leader for each subject who received psilocybin.

The experimenter held a briefing meeting with the leaders. Each leader received a printed summary of his purpose and functions, and a protocol of the experimental procedure. Stress was placed on a supportive but non-directive role so that each subject's experience could have its own development. The leaders were purposely not told the characteristics of the typology of mysticism and were not shown the post-drug questionnaire or any of the data already collected on the subjects. Neither the leaders nor the

experimenter knew which subject would receive psilocybin. The experimental procedure was discussed in detail so that the leaders could be of maximal help in making the experiment run smoothly. Emphasis was placed on creation of an atmosphere which would not create feelings of manipulation, suspicion, or psychotic terror in the subjects. Strategies for handling such disturbing reactions were reviewed (e.g., taking the subject out of the chapel and reassuring him). Leaders were instructed not to interfere unless a subject was having obvious difficulty. Group assignments were made, and the procedure and purpose of the group meetings to be held by the leaders with their groups before the experiment were discussed.

Each group of subjects met with its two leaders for two hours on one occasion two to five days before Good Friday. The meeting was held in the dormitory room of one of the group members. The purpose of this meeting was to develop group spirit and to prepare subjects for as positive and meaningful an experience as possible. Group members had an opportunity to become acquainted and feel at ease with their leaders. Subjects were encouraged to let themselves go into unexplored realms of experience during the actual experiment and not to try to fight the effects of the drug,

even if the experience became very unusual or frightening. The method of reaction to the experience was suggested, rather than specific content either positive or negative. The point was made that each person's drug experience was to be uniquely his own and could not be predicted with accuracy at the present state of knowledge. Possible physical symptoms such as nausea and vomiting were discussed as unwanted side effects which could be minimized or not experienced at all if the subjects were reassured that they were in good hands and that the experience, no matter how unusual, would be temporary. Personal preparation and the plan of the day on Good Friday were reviewed and discussed so that everyone was clear on the expected procedure. Subjects were instructed to eat an early, very light, non-fat breakfast. Each group selected its own meeting place on campus for the morning of the experiment.

Experimental Procedure

Drug Preparation

Double-blind technique was employed. The subjects had been told that some of them would receive psilocybin and some would be controls, but they did not know that two of the subjects in each group would receive psilocybin, nor that the other two members, as controls, would receive a control

substance, nicotinic acid. They had been told that one of their two leaders would receive a small dose of psilocybin. This was meant to add to the reassurance of the subjects and to aid group spirit and rapport with the leaders. Each pair of leaders knew that one of them and two of their group would receive psilocybin, but they were not told the nature of the control substance. Thirty capsules which were identical in outward appearance were prepared, eight hours before the experiment. Ten contained 30 mg. of psilocybin; five contained 15 mg. of psilocybin; and fifteen contained 200 mg. of nicotinic acid. Powdered sugar was used to fill any unused space in the capsules. Each capsule was sealed in an unmarked envelope, and the envelopes were kept in three groups, according to contents.

The list of paired subjects was given to a helper who did not participate in the experiment and did not know any of the subjects or leaders. He flipped a coin for each pair to designate which would be experimental and which would be control. He then checked the group lists to make sure that the experimental and control subjects were divided evenly in each group. In one instance, the coin-flip determination was arbitrarily changed to maintain this balance. A similar procedure was followed for the five pairs of

leaders. The helper then wrote the names on the appropriate envelopes, and the coding key was sealed in an envelope and locked in a drawer.

Protocol During the Experimental Day

The actual experiment in which psilocybin was administered to half of the participants occupied most of the day on Good Friday. A church in Boston provided exclusive use of a small prayer chapel which was connected by a hallway to three adjoining rooms in the basement. This chapel was large enough for fifty people. Two of the rooms were large, and the other was considerably smaller, but large enough to accommodate easily a group of six persons. The rooms were comfortably furnished with sofas and chairs. The "live" Good Friday service in the main sanctuary of the church upstairs was transmitted through a high-fidelity amplifier to speakers in the private chapel and two large rooms. Lighting in the chapel was controlled by a rheostat which was set for dim light. Candles were lit on the altar on either side of a golden cross. There were three stained-glass windows behind the altar. No incense was used. Exits from the experimental area were locked, except for one which could not be. A helper who was a clergyman was stationed

there throughout the experiment to keep all participants in the experimental area and to prevent any persons not connected with the experiment from entering. Signs were placed on the outside of all other exits to indicate that a private worship service was in progress.

On the morning of the experiment, the two leaders met the four members of their group at their own predetermined meeting place at about 9:30 AM and went as a group by automobile to the church. By 10:05 AM, all groups had assembled in the church basement. Separate areas were assigned to each of the five groups so that one group met in each end of the two large rooms, and the fifth group in the smaller room. Participants helped to arrange the sofas and chairs into small circles with a tape recorder in the middle of each area. An opportunity was given for each person to become familiar with the physical surroundings of the experimental area, i.e., the chapel, the three rooms opening off the hallway, and the toilet facilities which also were entered from the hall. A brief meeting was held for everyone in one of the large rooms and to enable all student participants to meet the other group leaders whom they had not seen before. The minister who was to "stand guard" at the unlocked

entrance was introduced and his function explained. The participants were encouraged to stay together in the chapel for the entire service, but were given the freedom to leave if they had to go to the toilet or if they felt restless. The understanding was made clear that everyone was to stay in the experimental area during the course of the day. An effort was made to allay any fearful tension and to encourage serious expectancy.

After this meeting, each group of four student volunteers and two leaders sat together in their assigned area. Shortly thereafter, at 10:30 AM, the envelopes were distributed to the group by the experimenter who watched as the capsules were swallowed with a small amount of water. For the next eighty minutes until it was time to move into the chapel, silence was observed in each group. Each individual was free to read, meditate or pray. The leaders were ready to help anyone who became frightened or experienced disturbing physical symptoms.

At 11:45 AM, a bell was sounded as the pre-arranged signal for all the group to move into the chapel. Each group sat together with their leaders. The minister welcomed everyone in the chapel before he went upstairs to start the service. An organ prelude was heard before the actual ser-

vice started.

The experimenter remained outside the chapel in the hall to be of assistance in case of emergencies. Oral and injectible thorazine were available as well as other standard emergency drugs. A written record was taken of the subjects and leaders who left the service from time to time to go to the toilet or to one of the other rooms. One of the large rooms was reserved for subjects who wanted to talk and the other for those who wanted to sit in silence. Each person who came out of the service was encouraged but not coerced to return. Most of the subjects remained in the chapel for the entire service.

At 2:30 PM when the service was over, the groups remained in the chapel. One leader went with one subject at a time back to the original group area where each subject described his experience into a tape recorder. After recording, each subject was escorted back to his group in the chapel. The other leader stayed with the group in the chapel during the recording.

At about 3 PM, after all the recordings were made, each group assembled in its area for a discussion and sharing of the experience. Fruit and juice were provided. All this conversation was recorded on tape.

By 4:30 PM, all subjects were judged recovered enough to leave. Each group was dismissed as a unit, and went out to eat together. After the meal, the leaders took each person home.

Collection of Data after the Experiment

The next morning, each subject called the experimenter to report his physical and psychological condition and how he had slept. Mental alertness and clarity, and presence or absence of headache, nausea, or dizziness were specifically elicited. The schedule of post-session interviews was arranged at this time.

In the days following the experiment, each subject wrote a description of his experience and mailed it to the experimenter. During the two-hour post-session interview, a 147-item questionnaire⁷ was completed by each participant in about 30 minutes before any discussion. Certain items were based on the categories of the typology of mystical experience. Other items measured physical symptoms or unpleasant reactions such as fear, terror, anxiety, loneliness, depression, or hatred. The person was asked to evaluate the degree to which he experienced each item on a scale from 0-4:

⁷See appendix C for a sample of this post-drug questionnaire.

- 0 - None, or not experienced at all.
- 1 - Experienced so slightly as to be doubtful.
- 2 - Experienced slightly
- 3 - Experienced moderately.
- 4 - Experienced strongly.

It was explained that the highest rating ("4" or "strong") was for extreme experiences. For an item to be scored as "4", the experience had to equal or excel. an experience which the participant considered strong, either in comparison to his personal past experiences or to what he would regard as strong in terms of each particular item. The written description and post-drug questionnaire were used as the basis for the ensuing interview. Unclear points were discussed. A tape-recording was made of each interview. A summary of each person's experience and reaction to Good Friday was written by the experimenter.

Six months after the experiment another two-hour period was devoted to each subject. A follow-up questionnaire which consisted of three parts was first completed.⁸ Part I was open-ended. The participant was asked to list any changes which he felt were a result of his Good Friday experience, and to rate the degree of benefit or harm of each change. Part II (52 items) was a condensed and some-

⁸See appendix D for a sample of this questionnaire.

what more explicit repetition of items from the post-session questionnaire. Part III (93 items) was designed to measure specific changes in attitude and behavior, both positive and negative. The same rating scale as in the post-session questionnaire was employed in all three parts except that the scale used was from 0-5. A distinction could be made in the highest rating as to whether this experience merely equalled what had previously been thought to be a strong experience ("4"), or excelled it ("5"). Thus "4" and "5" are more precise ratings of what would have been scored as "4" on the 0-4 scale of the post-session questionnaire. In the one-and-one-half-hour interview which followed, most emphasis was placed on a discussion of the nature of any changes which had been indicated. With the aid of Part II, the phenomena of the main categories of the typology were quickly reviewed. The interviewer took a more active role in this part than in the post-session interview. Categories were described in phenomenological terms, and the subject was asked to compare his Good Friday experience to these categories. The response to Part III was used as a basis for fuller description during the interview of any changes in attitude or behavior which had persisted for the six months since Good Friday.

Subjects were also invited to express their feelings about the whole experiment and asked how they would have preferred the set and setting to have been structured. The whole interview was taped. After each interview the experimenter wrote a summary of each subject's experience compared to the mystical typology. Subjects were not told whether they were experimentals or controls. The interviewer himself did not know until after completion of the content analysis of the individual writeups.

The accounts written within a few days after the Good Friday experience and those written six months later (Part I of the follow-up questionnaires) were content analyzed for phenomenological evidence of the categories from the typology of mysticism. The content analysis proceeded in three steps. First, the categories of this typology were described by the experimenter in the form of a training manual for judges who were unsophisticated in both mysticism and psychological experimentation. The description attempted neither to use theological concepts nor to identify the experience as "religious"; it sought instead to describe the phenomenology of the categories of mysticism without naming them as such. These phenomena were described in levels of completeness or intensity, corresponding to the

following 0-3 scale for each category:

- 3 - Experienced to a strong degree
- 2 - Experienced to a moderate degree
- 1 - Experienced to a slight degree
- 0 - None or not experienced at all

The judges were instructed to read the accounts and make a qualitative rating for the highest level attained in each experience for each category or sub-category.²

Second, the category description and rating system were pre-tested on five judges. Originally the plan was to score the content both for qualitative intensity or completeness, and for quantitative number of mentions of each level for each category. The results of this pre-test showed that the category descriptions were in general satisfactory but that there were too many confusing sub-categories. Also, the overlapping nature of the categories made quantitative scoring quite confusing, because the categories were imposed a priori on the accounts, rather than derived from the accounts. New, non-overlapping categories were not devised because a major purpose of the content analysis was to determine whether or not phenomena which would correspond to

²See appendix C for the training manual and the judges' score sheet.

the a priori categories were present before the administration of the questionnaire, which could have suggested such phenomena.

Third, the scoring of the categories was revised to a more simple method, and an additional three final judges were trained. These judges were former school teachers who were now housewives and mothers. They did not know the nature or design of the experiment, or that drugs had been given. They were not told that there were experimental and control groups. They were asked to score twenty experiences reported by persons who had attended a Good Friday worship service. The categories were explained by the written description in the training manual, and sample accounts were scored and discussed. All three judges attended the same training session and so were exposed to the same supplemental oral explanation. They received mimeographed copies of the accounts (60 single-spaced typewritten pages) which had been edited to remove all mention of receiving a drug and which were coded to remove all names. A different code was used for the authors of each account and for mention of these same persons in other accounts so that one writer's impression of someone else's experience would not bias the judges.

The accounts were arranged in the same order for each

judge. Matched pairs of subjects were placed consecutively with the order of experimental followed by control, or vice versa, determined by a coin-flip. The judges were instructed to read all the accounts through at one sitting and then to proceed with the individual scoring of each account one by one. With enough time allowed to finish all the categories in any account which was started, each judge gave a qualitative score to each subcategory or category on a 0-3 scale, as described in the training manual. (Twenty-two items for each account, or 440 in all, were used.) The attempt to score each qualitative level of each item quantitatively was dropped. The scoring procedure was designed so that each account was read five times and different categories scored each time. The easiest categories (e.g., loss of time and space, and positive mood) were scored first. The more difficult categories were then scored after the account became more familiar through repeated readings. This technique was meant to reduce the chance that data would be missed. Each judge estimated that between 15 and 20 hours were required for the scoring of all 20 accounts. The Kendall Rank Correlation Coefficient (τ)¹⁰ was found by com-

¹⁰S. Siegel, Nonparametric Statistics for the Behavioral Sciences (New York: Mc-Graw-Hill, 1956), pp. 213-223.

paring ranks of the total scores given by each judge for the same subjects. The ranking procedure using the 0-3 scale was found to be quite reliable. The rankings of the judges' scores were found to have a high correlation with each other ($r=.80$; p less than .001) (See appendix E).

There was total initial agreement among all three judges in 68.7% of the 440 items. After each judge's score on each item had been recorded, the three judges held a meeting in which consensus was reached on the remaining 31.3% or 138 items on which there had not been initial total agreement. The experimenter organized the meeting and explained the procedure, but did not participate in the discussion. He only listened to the deliberations from another room to gain an impression of how well the judges had understood the category definitions. For most of the items discussed two of the three judges were already in agreement. Most of the disagreement had been due to one judge having missed a piece of datum rather than a basic misunderstanding of the categories. The ease with which they came to final consensus reflected the good level of initial agreement as well as an adequate understanding of the categories. These final scores were used in statistical calculations which compared the scores of experimental and control subjects.