

The Effect of Experimenter Differences and Subjects' Belief Level upon ESP Scores¹

KARLIS Osis AND DOUGLAS DEAN²

ABSTRACT: An ESP experiment with lecture audiences as subjects was performed by two experimenters, each conducting four test sessions. The subjects' attitude toward ESP was ascertained by means of an incomplete sentence questionnaire and by two self-rating scales.

The hypothesis (Schmeidler's) that ESP scores would be positively related to the subjects' belief in ESP was confirmed in one experimenter's data (statistically significant multiple regressions), but not in the data of the other experimenter. The hypothesis (Van de Castle's) that the within subject variance would be higher in the pooled conflict and disbeliever (goat) categories than in the combined believer (sheep) categories was confirmed with an F test, as used by Van de Castle, at the .05 level of confidence. A chi-square test using subjects as a unit was not significant. A multiple regression analysis was performed on within subject variance and nearly reached significance in the results of one experimenter, but not in those of the other.

Exploratory measures of the subjects' emotional arousal and belief in their own ESP were also employed. Possibilities are discussed for more complex experiments concerning ESP in relation to belief and attitudinal variables.

This experiment was performed when both authors were with the Parapsychology Foundation, Inc. Dr. Osis is now Director of Research at the American Society for Psychical Research and Mr. Dean is Research Associate at the Newark College of Engineering. —Ed.

INTRODUCTION

In this experiment an attempt was made to repeat the essential features of the investigations of Drs. Gertrude R. Schmeidler and Robert L. Van de Castle. We were not prospectors looking for radically new findings; we were trying to mine gold where others had reported it to be. We were especially interested in the research problems explored by both Schmeidler and Van de Castle as they relate to current trends in the psychology of perception, that is, to the "new look" school. The "new look" psychologists have become increasingly aware that motivations and drives mold

¹ Part of this report was delivered to the Eastern Psychological Association's 33rd Annual Meeting, August, 1962.

² The duties of the two authors were divided. Both had an equal share in collecting the data. D. D. participated in judging the incomplete sentence questionnaire and in some of the evaluations. The design of the experiment, most of the statistical work, and writing the paper were K.O.'s responsibility. We are grateful to Dr. R. L. Van de Castle for his assistance as a consultant and judge, to Dr. M. E. Turner for valuable help on statistical problems, and to Miss S. Van Stenburgh for assistance in all stages of the project.

The Effect of Experimenter Differences

159

perception (6). This emphasis in normal perception led parapsychologists to explore the same factors in extrasensory perception.

Dr. Schmeidler, who was a student of one of the "new look" psychologists, Dr. Gardner Murphy, has done the lion's share of the work concerning conative factors in ESP. She found that subjects' belief that ESP is either possible or impossible strongly affects their ESP scoring level (5). Her results, which show remarkable consistency throughout 20 years of research, indicate that believers (sheep) score higher in ESP tests than disbelievers (goats), and that disbelievers exhibit an avoidance pattern; that is, they score below chance expectation. The disbelievers' results are in line with so-called "negative perception" findings in general psychology.

In Schmeidler's work, this differential scoring according to belief categories applied only to well-adjusted subjects (as judged by the Munroe Inspection Technique), while the scoring of maladjusted subjects was less predictable (4). The ESP scores of persons with a strong interest in theoretical problems (as tested by the Allport-Vernon Study of Values) were more predictable than those of persons having little interest in theory (5).

The essential features of Schmeidler's experiments enjoyed a number of replications by other researchers (1, 2, 5, 7). Generally concurring results emerged in seven out of nine experiments (Table 1), but the need for an improved technique to ascertain subjects'

TABLE 1
SUMMARY OF ESP SCORES ACCORDING TO BELIEF CATEGORIES
IN NINE PUBLISHED ESP EXPERIMENTS*

Experimenters	No. of Runs (25 tr.)	Sheep Dev- iation	Avg. per Run	No. of Runs (25 tr.)	Conflict Dev- iation	Avg. per Run	No. of Runs (25 tr.)	Goats Dev- iation	Avg. per Run	Hypothesis Confirmed*
Schmeidler (Individual)	1055	+242	5.23	—	—	—	853	—116	4.86	+
Schmeidler (Group)	3985	+614	5.10	—	—	—	4050	—301	4.93	+
Bevan	116	+44	5.38	136	+66	5.57	120	+2	5.02	+
Caspar	324	+51	5.16	208	—63	4.70	52	—8	4.85	+
Eilbert	70	+31	5.58	115	+2	4.98	20	—2	4.90	+
Kahn	733	+42	5.06	95	+22	5.23	48	—9	4.81	+
Woodruff and Dale	3000	+8	5.00	—	—	—	2800	+157	5.05	—
Van de Castle	648	+41	5.06	528	+6	5.00	376	—14	4.96	+
Adcock and Quartermain	144	—38	4.74	—	—	—	162	—9	4.94	—

* This Table includes only those experiments using the 5-choice guessing method. There are also published experiments having sheep-goat categories using different numbers of choices, as well as free choice stimulus material, but these are not included because of the difficulty of comparison.

** Scores of sheep higher than those of goats.

beliefs was stressed by several authors (2, 8, 9). Thus, the scales on which belief is measured were expanded from two- to three-, four- and five-point scales by Bevan, Caspar, Eilbert, Kahn, Schmeidler (in her later work), Van de Castle and White, and

others. The criteria of belief included (a) belief in the existence of ESP, (b) belief in the adequacy of the particular ESP test, and (c) belief in personal ESP. A detailed discussion of the problem of separating subjects into attitudinal categories can be found in Schmeidler and McConnell (5, Ch. 6), Mangan (2), and Van de Castle (7).

Dr. Robert L. Van de Castle developed an incomplete sentence questionnaire (7) for determining attitude toward ESP, rather than relying on self-ratings or ratings by the experimenter. From the results of his pilot study he realized that he did not merely have a scale from more to less belief: the middle group seemed to represent a conflict between belief and disbelief which, of course, is a qualitatively different motivational situation. He postulated that the ESP scores of the conflict group would be more erratic (the same subjects scoring high on some runs and low on others) and would therefore give higher "within subject" variance³ than scores of subjects whose beliefs were more homogeneous. The disbelievers were grouped with the conflict subjects after Van de Castle found that they performed similarly in his pilot studies. He suggested that disbelievers also have a conflict between a conviction that "ESP is the bunk" and social obligations toward the experimenter. The conflict subjects are faced with an internal conflict relatively independent of the testing situation; the disbelievers' conflict is thought to arise solely in the experimental setting and is therefore situationally labile.

Van de Castle's predictions (9) were confirmed in six repetitions (7) of similarly designed experiments ($P \leq .01$). He also found certain consistent patterns in the levels of "between subjects" variance for the different belief categories. The strong believers (high sheep) and indecisives (conflict) had much higher variance than the moderate believers (sheep) and disbelievers (goats). Five out of six F ratios resulting from a comparison of between subjects variance for the belief groups were statistically significant at a level ranging from .05 to .001.

The main problems investigated in the present study are the relationship between ESP scoring level and attitude toward ESP (Schmeidler's findings) and the differences in variance levels within subject and between subjects (Van de Castle's findings). In regard to belief criteria, we found criterion (a), belief in the existence of ESP, to be the most thoroughly explored and effective. It was adopted for the present experiment. We also used (c), belief in personal ESP, as a purely exploratory measure. We discarded (b),

³ For a discussion of the application of the analysis of variance to this problem, see Van de Castle (7, pp. 56-58).

belief in the adequacy of the particular ESP test, because it seems to be a concept under which too many theoretical, technical, situational, and sociological variables can be subsumed.

SUBJECTS AND EXPERIMENTAL PROCEDURE

The 246 subjects were persons attending a series of eight lectures (with accompanying slides) devoted to various aspects of ESP research, given in 1959-60 by the authors.⁴ Those participating in three of the eight sessions were members of women's clubs, three were church groups, one an educational group, and one a freshman college class. Miss S. Van Steenburgh assisted in the experiments. Whereas Van de Castle talked to his subjects on the topic of ESP for only about five minutes, we found this an inadequate compensation for their patience during a long test and we increased our lecture time to 40 minutes. We did not discuss Van de Castle's work in the lectures, nor the sheep-goat work in general.

We used the Van de Castle-White method of testing ESP in groups (9). Each subject was given a concealed list of four columns of ESP card symbols (100 in all) in random order (Van de Castle-White used eight columns). The stimulus symbols were randomized using the million-digit compilation of the Rand Corporation, *A Million Random Digits*. The random order of the ESP symbols was different for each subject. Each list was enclosed in an opaque manila envelope between two sheets of black construction paper. Two record sheets with carbon paper inserted between them were stapled on the outside of the envelope exactly over the target list inside, letting the blank columns precisely match the columns underneath inside the envelope. The experimenters tested the opaqueness in lighting considerably above the normal. It was foolproof in the given conditions of lighting and supervision.

Immediately after the lecture the target envelopes were distributed to the subjects, together with a mimeographed instruction sheet. The experimenter read the instructions aloud and answered questions. The subjects were supervised by two persons. Upon completion of the ESP tests, the envelopes were collected and the questionnaires (see Appendix) ascertaining attitude toward ESP were distributed. They consisted of Van de Castle's incomplete sentence questionnaire (e.g., "I believe that hunches often ———.") and two five-point self-rating scales ranging from strong conviction

⁴ The experiment was originally planned with K.O. as the only experimenter. This plan had to be changed after he became ill for several months, during which time D. D. took over his lecture commitments. This change modified the design of the experiment by introducing a second experimenter as another variable.

about the existence of ESP (high sheep) to complete disbelief in it (extreme goat). Scale A rated the subject's belief in ESP in general and Scale B the subject's belief in his own personal ESP ability. The incomplete sentence questionnaire was administered right after the completion of the ESP test; then the subjects were asked to turn the page and fill in self-rating scales A and B.⁵ Scoring of the ESP tests was done by an assistant and rechecked by another person. The incomplete sentence questionnaires were independently judged by three persons: K.O., D.D., and Van de Castle. All the judging was done as a blind procedure, i.e., without the knowledge of the subjects' names or ESP test scores. An effort was made to follow as closely as possible Van de Castle's original checking procedure, which he has described as follows:

The method of evaluation was based upon an over-all clinical impression of the subject's attitude. All the individual subject's responses were first read over in order to form a preliminary impression and then read again in order to uncover any discrepancies between individual items and total impression or between individual items themselves. Subjects were eventually placed into one of four attitudinal categories: high sheep, sheep, conflict, or goat.

In order to be considered a high sheep, the subject had to display a consistent attitude of moderate belief on all items or else display a high level of belief on most of the items with the remaining items containing neutral expressions of opinion.

When a subject indicated moderate belief on most of the items with the remaining items containing neutral expressions of opinion, he was classified as a "regular" sheep.

Two different patterns of responses were included in the conflict category. If a subject indicated belief in ESP on some of the items while at the same time he indicated strong doubts and suspicion regarding ESP on other items, a rather obvious conflict situation was present and such a subject was labelled accordingly. The other type of response pattern which was considered as representing a subject who possessed a conflicting attitude toward ESP was not so apparent. This subject might also be called an evasive one, for he would refrain from making any decisive commitment throughout all ten items by using such camouflaging responses as "is interesting," "is strange," "is not well understood," etc. Other clues which were considered as evidence of conflict were excessive erasures or crossed-out responses, not completing items, stereotyped use of the same adjective or its antonym, completely irrelevant responses, or long

⁵ It was decided to ascertain the subjects' belief category after the ESP test in order to keep the dynamics of attitude factors active on a low level of conscious awareness. It was thought that if beliefs and attitudes were made conscious by stating them beforehand, they might lose some of their affective charge and force the subject to act out a more superficial and clearly conscious commitment.

involved explanations to items that would contain anywhere from ten to twenty words.

Subjects classified as goats were those who gave indications of a negative or disbelieving attitude toward ESP on all or most of the items, providing that in the latter case the remaining items were neutral in expression (7, pp. 46-47).

In five cases there was complete disagreement among the three judges. Since there was no opportunity for a judging conference to come to an agreement, the data of these five subjects were excluded from statistical evaluations relating to the incomplete sentence questionnaire.⁶ In some evaluations of the data we separated those subjects on whose questionnaires all three judges agreed from those where only two agreed.

The self-rating questionnaires were independently checked by two persons other than the authors.

RESULTS

The previous studies summarized in Table 1 led us to expect that any extrachance results occurring in the sheep or goat categories might be obscured in the pooled totals: in other words, the positive scores of believers might be cancelled out by the negative scores of disbelievers. As later analyses will show, this was indeed the case. The total deviation of the 183 subjects whose data were included in the analyses was very close to mean chance expectation ($CR = .68$, $P = .50$).

However, division of the data according to experimenters yielded interesting results. K.O.'s subjects scored 82.2 hits above mean chance expectation ($CR = 2.18$, $P = .03$), while D.D.'s subjects scored 45.4 hits below mean chance expectation ($CR = 1.17$, $P = .24$). The difference between scores of the two experimenters is associated with a chi square of 5.6, 1 d.f., $P = .02$. D.D.'s and K.O.'s subjects taken as a unit (2 by 2 table of subjects scoring above or at and below chance expectation) yielded a chi square of 3.90, 1 d.f., $P = .05$. It thus appears that the subjects responded quite differently to the lectures and testing procedures of K.O. and D.D., although both experimenters used the same lecture outline, slides, and test instructions. However, it was not our

⁶ We had to exclude from all evaluations the data of those subjects who failed to respond to the questionnaires and also of those who did not complete the ESP test, i.e., where more than eight of the 100 trials were missing. On this basis we eliminated 63 subjects; their scores were close to chance ($CR = .54$). The scores of four subjects missing from one to eight trials were rounded out on a 100-trial basis by adding the average per trial scores times the number of trials missing. This decreased the total deviation by .29, an obviously insignificant change.

intention to test the differences in sheer scoring levels; our primary concern was with the various kinds of interaction between the subjects' belief and ESP scoring level. Our problem, in a nutshell, was to relate three aspects of ESP results (average scores, within subject variance, and between subjects variance) to two experimental measures of belief in ESP and to one exploratory belief measure.

As stated before, this experiment was largely a repetition of the work of other investigators; we therefore had to use some of their evaluation methods in order to permit comparisons, e.g., analyses of variance as they were applied in Van de Castle's experiments (7). Our statistical consultant considered multiple regression analyses more appropriate and safer, e.g., not involving the assumption of homogeneity of variance; we therefore performed multiple regressions in order to obtain more reliable and adequate information about our data. Although we feel that the multiple regression analyses carry the main weight, they should be interpreted in context with the other evaluations. These various evaluation methods basically deal with the same trends; therefore the results, particularly the P values, are often not additive. Customary caution in interpreting P values of multiple measurements is necessary.

We will present our results in the following order: (1) ESP scores in relation to belief categories, (2) within subject variance and belief categories, and (3) between subjects variance and belief categories.

Incomplete Sentence Questionnaire

In accordance with the results of the incomplete sentence questionnaire, we classified the subjects of both experimenters as follows: high sheep (strong belief in ESP), sheep (belief in ESP), conflict (indications of conflict or evasiveness), and goat (disbelief in ESP). All three judges agreed on the classification of 47 of K.O.'s subjects and 53 of D.D.'s; two judges agreed on the classification of an additional 40 of K.O.'s subjects and 38 of D.D.'s.⁷ We decided to pool the three- and two-judge agreement classifications because of the small number of subjects in the high sheep and the goat classifications (Table 2). This seemed to be justified since the scores in the three- and two-judge classifications showed roughly the same pattern.

⁷ As stated above, the data of five subjects on whose classification all three judges disagreed were excluded from statistical evaluations relating to the incomplete sentence questionnaire.

The patterns of average scores in the four belief classifications are quite different for the two experimenters (Table 2 and Figure 1). K.O.'s high sheep scored highest; then the sheep and conflict groups; and the goats lowest, whose scores were considerably below mean chance expectation. The average scores of D.D.'s subjects give a slight curvilinear pattern. Curiously enough, the scores of the high sheep were below chance expectation; those of sheep and conflict subjects at the chance level; while goats scored far below chance expectation.

TABLE 2
AVERAGE ESP SCORES ACCORDING TO BELIEF CATEGORIES

Belief Determination	Experimenter	High Sheep		Sheep		Conflict		Goats		Extreme Goats	
		No. Ss	Avg. per Run	No. Ss	Avg. per Run	No. Ss	Avg. per Run	No. Ss	Avg. per Run	No. Ss	Avg. per Run
Incomplete Sentence Questionnaire	KO	9	5.69	25	5.31	47	5.21	6	4.58	—	—
	DD	11	4.77	27	5.04	40	5.01	13	4.42	—	—
Self-rating Scale A	KO	24	5.65	42	5.20	8	4.87	5	5.05	4	4.62
	DD	27	4.76	37	5.08	16	5.09	4	5.19	6	4.12
Self-rating Scale B	KO	19	5.63	37	5.07	16	5.55	12	5.12	2	4.87
	DD	18	4.85	37	5.05	9	4.69	16	5.17	9	4.47

In order to assess the consistency of these scoring trends, we classified subjects into those who scored above mean chance expectation and those who scored at or below. There was a different distribution between K.O.'s and D.D.'s subjects who scored above chance expectation and those who scored at or below in the pooled high sheep and sheep category, with at least two judges agreeing ($\chi^2 = 4.72$, 1 d.f., $P = .03$), but a similar distribution in the conflict and goat categories. Twenty-three subjects in K.O.'s combined sheep category scored above chance expectation and 11 at or below. One goat scored above chance expectation and five at or below. This difference in proportions as assessed by the Fisher exact probability test was significant at the three per cent level. The same kind of test of D.D.'s data yielded a non-significant result.

In correspondence Van de Castle commented that it would be interesting to see the results of the subjects where all three judges agreed that he was either a high sheep or a sheep; in other words, whether he was a believer in ESP, regardless of the intensity of the belief. Accordingly, such an analysis was performed. K.O.'s eligible combined sheep (25) had an average score of 5.42. D.D.'s eligible combined sheep (32) had an average score of 4.97, which is very close to mean chance expectation. In both experimenters' data combined sheep scored higher than goats: for K.O., 5.42 versus 4.58; for D.D., 4.97 versus 4.42.

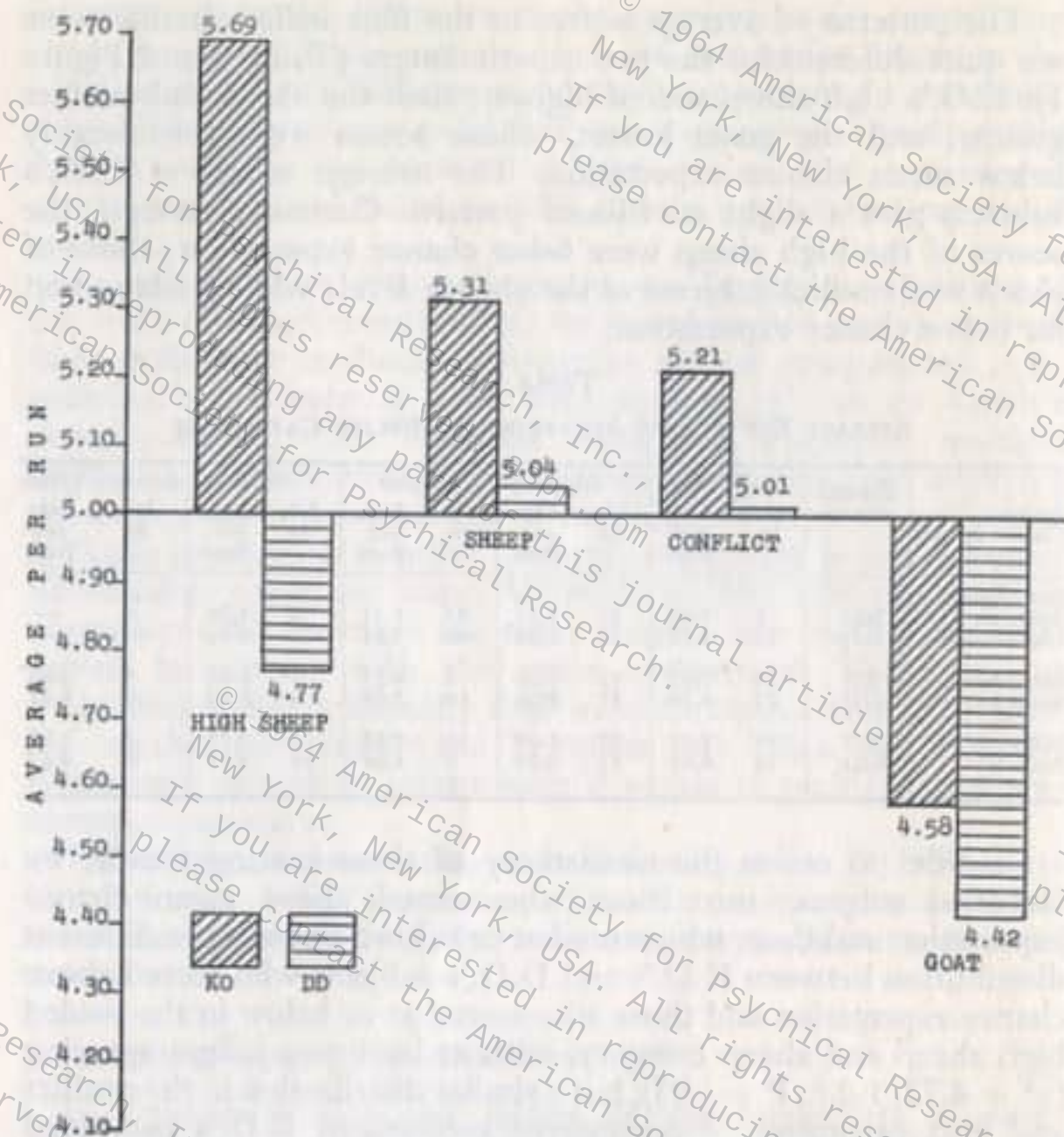


FIG. 1. Comparison of scores between the two experimenters of subjects in different belief categories, as determined by incomplete sentence questionnaire, at least two judges agreeing.

Self-Rating Scale A

Self-rating scale A separated the subjects in an even more dramatic way (Table 2 and Figure 2). Distribution of ESP scores obtained by this measure was similar to that using the incomplete sentence questionnaire classifications. The scores of K.O.'s subjects were again in a nearly linear relationship to the belief categories: the high sheep had a 5.65 average score; the sheep 5.20; conflict group 4.87; goats 5.05, and the extreme goats (lowest self-rating category) 4.62. The scores of D.D.'s subjects also gave roughly the same picture as when separated by the incomplete

sentence questionnaire. We again separated the subjects who scored above mean chance expectation from those who scored at or below. A significant difference between experimenters emerged from the combined sheep categories ($\chi^2 = 6.99$, 1 d.f., $P = .014$). No significant differences were registered in the combined remaining categories.

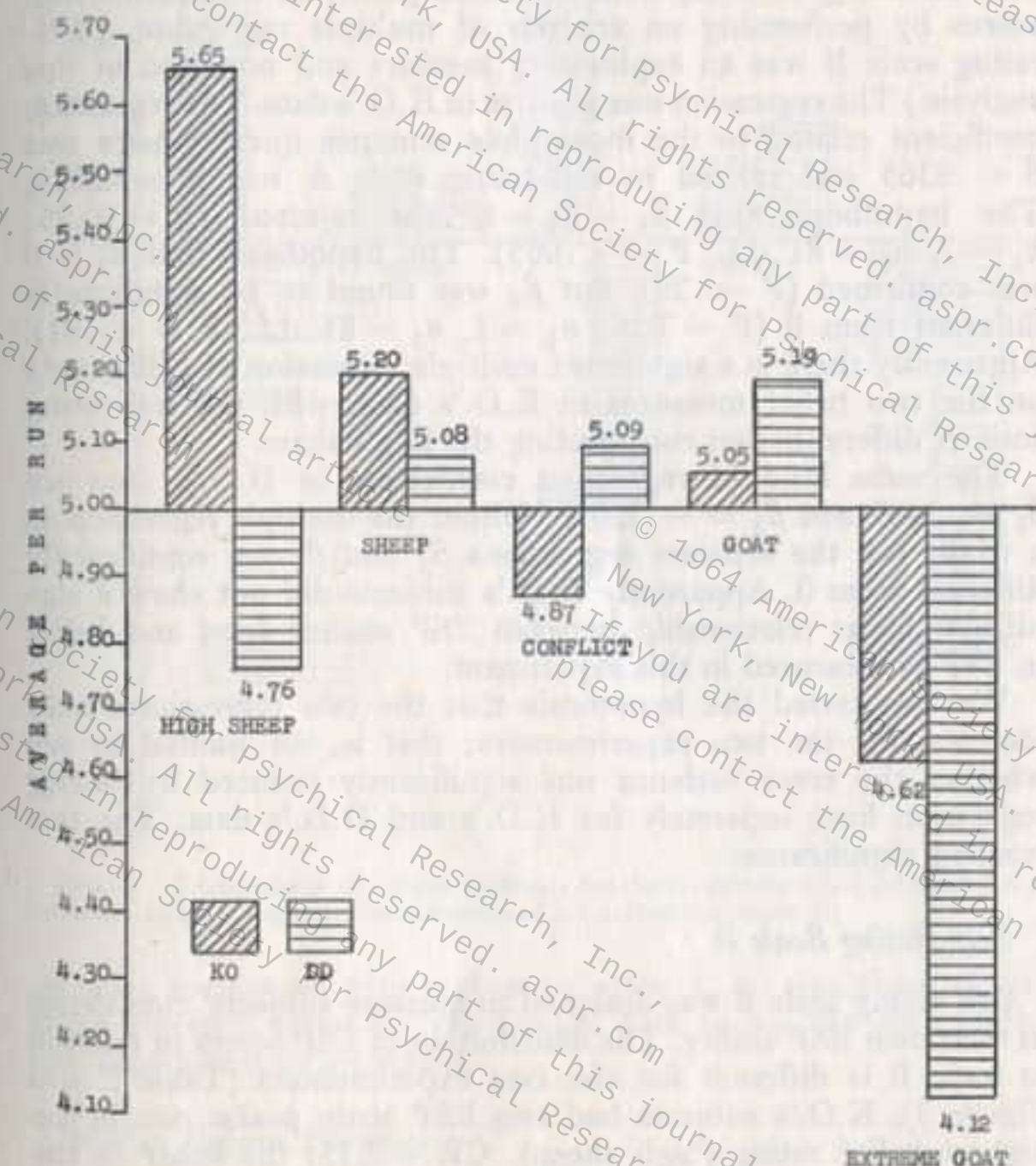


FIG. 2. Comparison of scores between the two experimenters of subjects in different belief categories, as determined by self-rating scale A.

K.O.'s subjects scoring above or at and below chance expectation in the combined belief categories 5 plus 4 versus 3 plus 2 plus 1 were distributed in a definitely non-chance pattern ($\chi^2 = 9.53$, 1 d.f., $P = .002$). A similar chi-square test of D.D.'s data

gave $\chi^2 = .64$, which is insignificant. It appears that separating subjects by self-rating scale A reveals slightly more differentiated scoring patterns than separating them by the incomplete sentence questionnaire. Figures 1 and 2 accentuate the experimenter differences.

These two belief measures (incomplete sentence questionnaire and self-rating scale A) were evaluated jointly in relation to ESP scores by performing an analysis of multiple regression. (Self-rating scale B was an exploratory measure and not used in this analysis.) The regression was positive in K.O.'s data. The regression coefficient related to the incomplete sentence questionnaire was $\beta = .8365$ and related to self-rating scale A was $\beta = 1.2024$. The hypothesis that $\beta_1 = \beta_2 = 0$ was rejected ($F = 10.26$; $n_1 = 2$, $n_2 = 81$ d.f., $P = < .005$). The hypothesis that $\beta_1 = 0$ was confirmed ($F = .29$), but β_2 was found to be significantly different from 0 ($F = 7.56$; $n_1 = 1$, $n_2 = 81$ d.f., $P = < .01$). Apparently there is a significant multiple regression of ESP scores on the two belief measures in K.O.'s data, with the self-rating scale A differentiation contributing the lion's share.

The same kind of regression coefficients in D.D.'s data are $\beta_1 = .2409$ and $\beta_2 = -.7201$. Neither the multiple regression as a whole nor the separate regressions β_1 and β_2 are significantly different from 0. Apparently D.D.'s subjects did not show a significant linear relationship between ESP scoring level and belief in ESP as measured in this experiment.

We also tested the hypothesis that the two regressions were identical for the two experimenters; that is, we wanted to see whether the error variance was significantly reduced by fitting regression lines separately for K.O.'s and D.D.'s data. The test reached significance.

Self-Rating Scale B

Self-rating scale B was designed to measure subjects' confidence in their own ESP ability. The distribution of ESP scores in relation to scale B is different for the two experimenters (Table 2 and Figure 3). K.O.'s subjects had two ESP score peaks: one in the highest belief rating (high sheep), $CR = 2.75$; the other in the middle rating (conflict group), $CR = 2.21$. If these measures indicate anything more than a chance distribution, as they well might, we may be dealing with a different dimension of belief probably involved with the personality factor of self-confidence. When we separated the combined sheep from the conflict group and combined goats according to scores above and at or below mean chance expectation and performed chi-square tests similar

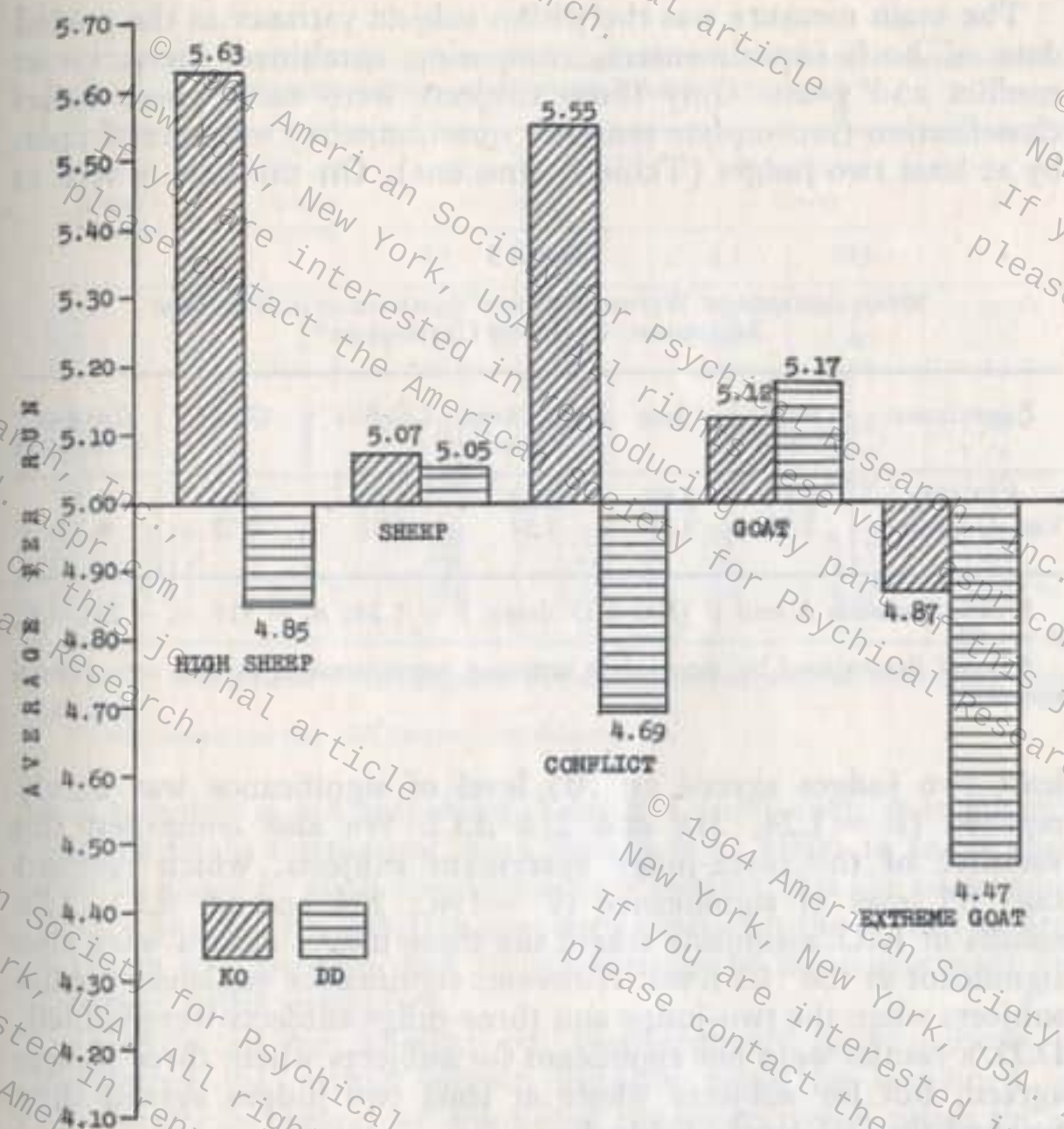


FIG. 3. Comparison of scores between the two experimenters of subjects in different belief categories, as determined by self-rating scale B.

to those performed with self-rating scale A, no significant differences emerged either for the pooled data or for the data of the two experimenters separated.

Within Subject Variance

Van de Castle compared scoring variance (*note*: not the means) between the eight runs of 25 trials by obtaining an individual sum of squares for each subject. According to Van de Castle, "It involves taking each run score, finding the difference between it and the subject's average run score, then squaring this difference and summing all these differences" (7, p. 53). The same technique was followed in the present experiment in four-run blocks.

The main measure was the within subject variance in the pooled data of both experimenters, comparing combined sheep *versus* conflict and goats. Only those subjects were used whose belief classification (incomplete sentence questionnaire) was agreed upon by at least two judges (Table 3, line one). On the data where at

TABLE 3

MEAN SQUARES OF WITHIN SUBJECT VARIANCE OF ESP SCORES
ACCORDING TO BELIEF CATEGORIES*

Experiment	High Sheep	Sheep	High Sheep + Sheep A	Conflict	Goats	Conflict + Goats B
KO-DD	3.14	3.67	3.52	4.61	3.25	4.37
Van de Castle	3.78	3.39	3.54	4.81	4.71	4.77

F ratio between A and B (KO-DD data): $F = 1.24$; $n_1 = 318$, $n_2 = 216$ d.f., $P = .05$

* Belief determined by incomplete sentence questionnaire, at least two judges agreeing.

least two judges agreed an .05 level of significance was barely reached ($F = 1.24$, 318 and 216 d.f.). We also computed the variance of the three-judge agreement subjects, which reached the .05 level of significance ($F = 1.40$, 204 and 96 d.f.). The results of K.O.'s subjects where the three judges agreed were also significant at the .05 level. However, significance vanished for his subjects when the two-judge and three-judge subjects were pooled. D.D.'s results were not significant for subjects where three judges agreed, but for subjects where at least two judges agreed they reached the .05 level (Table 4).

Van de Castle suggested that we also work out the within subject variance on data where all three judges agreed that the subject should be classed as some kind of sheep, regardless of intensity of belief, or should be placed in the combined conflict and goat group, without distinction between them. The within subject variance of subjects separated in this manner did not reach significance because of a slightly smaller F value and fewer degrees of freedom ($F = 1.21$, 234 and 171 d.f.).

We compared the average within subject variance per run between the two experimenters and Van de Castle's original data (Figure 4.) This figure suggests that D.D.'s subjects followed Van de Castle's original pattern more closely than did K.O.'s. In the goat classification neither D.D.'s nor K.O.'s subjects followed Van de Castle's pattern, but there were only 19 subjects classed as goats in our group. This lack of similarity in scoring level in the

TABLE 4

MEAN SQUARES OF WITHIN SUBJECT VARIANCE OF ESP SCORES ACCORDING
TO EXPERIMENTERS AND POOLED BELIEF CATEGORIES*

Experimenter	No. of Judges Agreed	High Sheep + Sheep		Conflict + Goats		F
		d.f.	MS	d.f.	MS	
KO	3	42	2.87	99	4.58	1.60**
KO	2	60	4.27	60	3.40	1.26
KO	3 + 2	102	3.69	159	4.14	1.12
DD	3	54	3.77	105	4.84	1.28
DD	2	60	3.05	34	4.14	1.36
DD	3 + 2	114	3.27	159	4.60	1.36**
KO + DD	3	96	3.38	204	4.72	1.40**
KO + DD	2	120	3.66	114	3.75	1.02
KO + DD	3 + 2	216	3.52	318	4.37	1.24**

* Belief determined by incomplete sentence questionnaire, at least two judges agreeing.

** Significant at the .05 level of confidence.

goat category is not surprising. Van de Castle's own goat subjects had variances fluctuating more widely from series to series than did the sheep or conflict subjects. It is evident that the patterns of K.O.'s and D.D.'s subjects are more similar in the within subject variance comparisons than in comparisons of mean scores in the different belief categories.

In order to see whether our comparatively slight success in separating the within subject variance could be explained by differences in judges, we computed the respective F ratios for all three judges. They were almost identical: Van de Castle, $F = 1.22$; D.D., $F = 1.18$; K.O., $F = 1.22$. Apparently the combination of judges increased the F ratio slightly over the single judge approach. It is clear that the present experimenters were able to use Van de Castle's incomplete sentence questionnaire technique with an efficiency equal to that of the original experimenter.

The F ratio measure Van de Castle used was marginally significant in our data. However, he saw a possible loophole in the measure and used an alternative evaluation as follows:

As was mentioned earlier, however, the mean can be heavily influenced by a few atypically high or low scores, and since this F ratio resulted from a ratio of two mean sums of squares, it is possible that this significant F value is due to a few atypically scoring subjects. In order to check upon the consistency of these scoring patterns, the chi-square technique was employed. The over-all mean sum of

squares of all 194 subjects was computed and found to be 29.787, and the number of subjects having sums of squares scores above and below this over-all mean was tallied . . . A 2 x 2 table was set up containing the number of subjects scoring above and below the over-all mean for the combined groups . . . (7, p. 55).

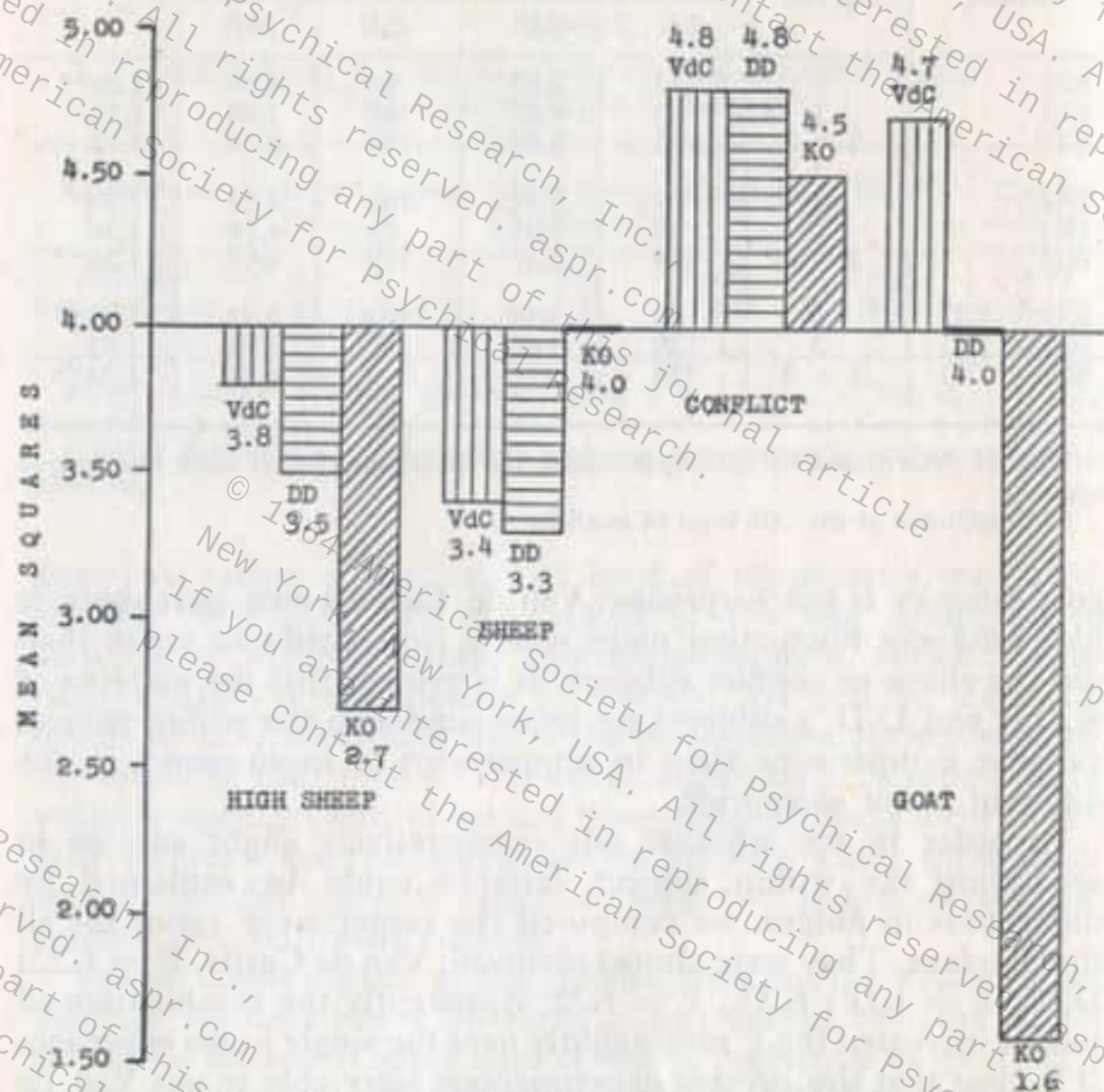


FIG. 4. Within subject variance in Van de Castle's experiment and in the present series, belief determined by incomplete sentence questionnaire, at least two judges agreeing.

In the present experiment we computed a 2 x 2 table of subjects whose variance was above and at or below the mean for the experiment, with high sheep and sheep combined, and conflicts and goats combined. The results were very close to chance expectation ($\chi^2 = .17$, 1 d.f.). Therefore, the F ratio measure of the within subject variance must be interpreted with great caution. However, Van de Castle's argument does not apply to our regression analyses described later.

Within subject variance was also computed on subjects separated according to self-rating scale A (Table 5). We combined the data of high sheep and sheep, and those of the conflict and goat categories. The results of both K.O.'s and D.D.'s subjects when pooled were very close to chance expectation (mean square for combined

TABLE 5
MEAN SQUARES OF WITHIN SUBJECT VARIANCE OF ESP SCORES
ACCORDING TO BELIEF CATEGORIES*

Experimenter	(1) High Sheep	(2) Sheep	(3) Conflict	(4) Goats	(5) Extreme Goats	F ratio between 1 + 2 3 + 4 + 5
K.O.	4.59	4.17	2.83	2.78	2.79	1.54**
DD	2.76	4.51	4.75	5.27	4.29	1.25

* Belief determined by self-rating scale A.

** Significant at the .05 level of confidence.

sheep, 4.05; mean square for combined conflict-goat group, 3.97). However, when we separated the subjects according to experimenters, the variance of D.D.'s subjects roughly followed Van de Castle's pattern, with the mean squares in the five belief categories being 2.76, 4.51, 4.75, 5.27, and 4.29 (not statistically significant according to Van de Castle's method of analysis of variance). K.O.'s subjects revealed just the opposite pattern from the one obtained with the incomplete sentence questionnaire and was significant at the .05 level. The mean squares in the five belief categories—4.59, 4.17, 2.83, 2.78, and 2.79—show more variance with more belief, a pattern which is contrary to our expectations as based on Van de Castle's findings. This reversal of the variance pattern in K.O.'s subjects might reflect subtle differences in the assessment of belief: the incomplete sentence questionnaire and self-rating scale A might have measured different dimensions of belief. Because of the level of statistical certainty ($P = .05$), the chance explanation of this unpredicted trend is not excluded.

We evaluated the within subject variance in joint consideration with the incomplete sentence questionnaire results and the self-rating scale A scores by performing multiple regression analyses. The variance of each subject's four runs in relation to his own mean was the dependent variable Y . The incomplete sentence score (where at least two judges agreed) was the independent variable X_1 , and the self-rating scale A score was the independent variable X_2 .

The regressions went in opposite directions in K.O.'s data. The regression coefficient related to the incomplete sentence questionnaire was $\beta_1 = -1.9665$; but the coefficient related to self-rating scale A was $\beta_2 = +2.4500$.

An F test of the hypothesis that $\beta_1 = \beta_2 = 0$ and that $\beta_1 = 0$ gave non-significant results. The hypothesis that $\beta_2 = 0$ was rejected at the .05 confidence level ($F = 3.96$; $n_1 = 1$, $n_2 = 81$ d.f.). Apparently the joint consideration of the two belief measures in K.O.'s data did not conform to the patterns predicted on the basis of Van de Castle's findings.

In D.D.'s data both regressions went in the predicted direction. The regression coefficient related to the incomplete sentence questionnaire was $\beta_1 = -1.912$ and related to self-rating scale A was $\beta_2 = -2.3270$. As in K.O.'s data the β_2 coefficient was numerically larger than β_1 , but unlike K.O.'s data it went in the direction expected from Van de Castle's hypothesis.

An F test of the hypothesis that $\beta_1 = \beta_2 = 0$ came close to the .05 confidence level ($F = 3.00$; $n_1 = 2$, $n_2 = 85$ d.f., when $F_{.05} = 3.11$). The hypothesis that $\beta_1 = 0$ was accepted ($F = .02$). The F test of the hypothesis that $\beta_2 = 0$ again gave an F value close to the .05 level of confidence ($F = 3.86$; $n_1 = 1$, $n_2 = 85$ d.f., when $F_{.05} = 3.96$). On the whole, the multiple regressions performed on D.D.'s data gave results somewhat similar to Van de Castle's findings.

As in the case of multiple regressions on ESP scores and belief variables, we also tested the hypothesis that the two regressions were identical for the two experimenters; that is, we wanted to see whether the error variance was significantly reduced by fitting regression lines separately for K.O.'s and D.D.'s data. The test reached significance at the .05 confidence level ($F = 2.80$; $n_1 = 3$, $n_2 = 166$ d.f.).

Between Subjects Variance

Van de Castle found certain differential patterns of between subjects variance within the sheep-goat classifications. In this experiment it was a surprise to find that the between subjects variance was significantly different in K.O.'s and D.D.'s subjects, with the variance of D.D.'s subjects about one-third higher than K.O.'s ($F = 1.60$; 87 and 83 d.f.; $P = .02$). Because of the apparent dissimilarity of variance, the between subjects variance was evaluated separately for the subjects of both experimenters (TABLE 6).^a We also computed six F ratios for each experimenter's

^a Every Mean Square (MS) of Table 6 (which represents variability in each belief category: high sheep, sheep, conflict, and goat) is compared with the MS in

subjects (TABLE 6A). One of the F values in each experimenter's data reached the .05 level of significance which is quite close to chance expectation.^b

TABLE 6
ANALYSIS OF VARIANCE TABLE FOR EVALUATING DIFFERENCES BETWEEN SUBJECTS' TOTAL SCORES WITHIN DIFFERENT BELIEF CATEGORIES*

Category	KO			DD		
	SS	d.f.	MS	SS	d.f.	MS
High Sheep	111.56	8	13.94	220.91	10	22.09
Sheep	312.53	24	13.02	285.41	26	10.98
Conflict	447.71	46	9.73	858.62	39	22.02
Goats	25.33	5	5.07	142.62	12	11.88
Totals:	897.13	83	10.81	1507.56	87	17.33

KO versus DD: $F = 1.60$, $P = .02$

* Belief determined by incomplete sentence questionnaire, at least two judges agreeing.

TABLE 6A
F RATIOS BETWEEN BELIEF CATEGORIES SHOWN IN TABLE 6

Category	KO			DD		
	Sheep	Conflict	Goats	Sheep	Conflict	Goats
High Sheep	1.07	1.43	2.75	2.01	1.00	1.86
Sheep		1.34	2.57*		2.01*	1.08
Conflict			1.92			1.85

* $P = .05$

We wanted to compare our between subjects variance with Van de Castle's results, but this could not be done accurately since he had twice as many trials for each subject. We found that his

every other belief category. Thus we obtained six comparisons for each experimenter's data from which we computed the same number of F ratios presented in Table 6A.

^b In a comment on the manuscript Van de Castle suggested that there appears to be a greater similarity between his and D.D.'s pattern if the rank order of magnitude of the variances is taken into account: "If you have four scores (1, 2, 3, 4), there are 24 possible different arrangements that these numbers can appear in; and out of 24 such possible permutations D.D. and I agreed exactly in rank position on the order of between subjects variance with both of us having the following order: high sheep (1), conflict (2), goat (3), and sheep (4). This result by itself would be significant beyond the .05 level."

between subjects mean squares were about 2.7 times higher than ours; therefore we divided his values by 2.7. These adjusted values, of course, form a basis for only a very crude comparison (Figure 5).

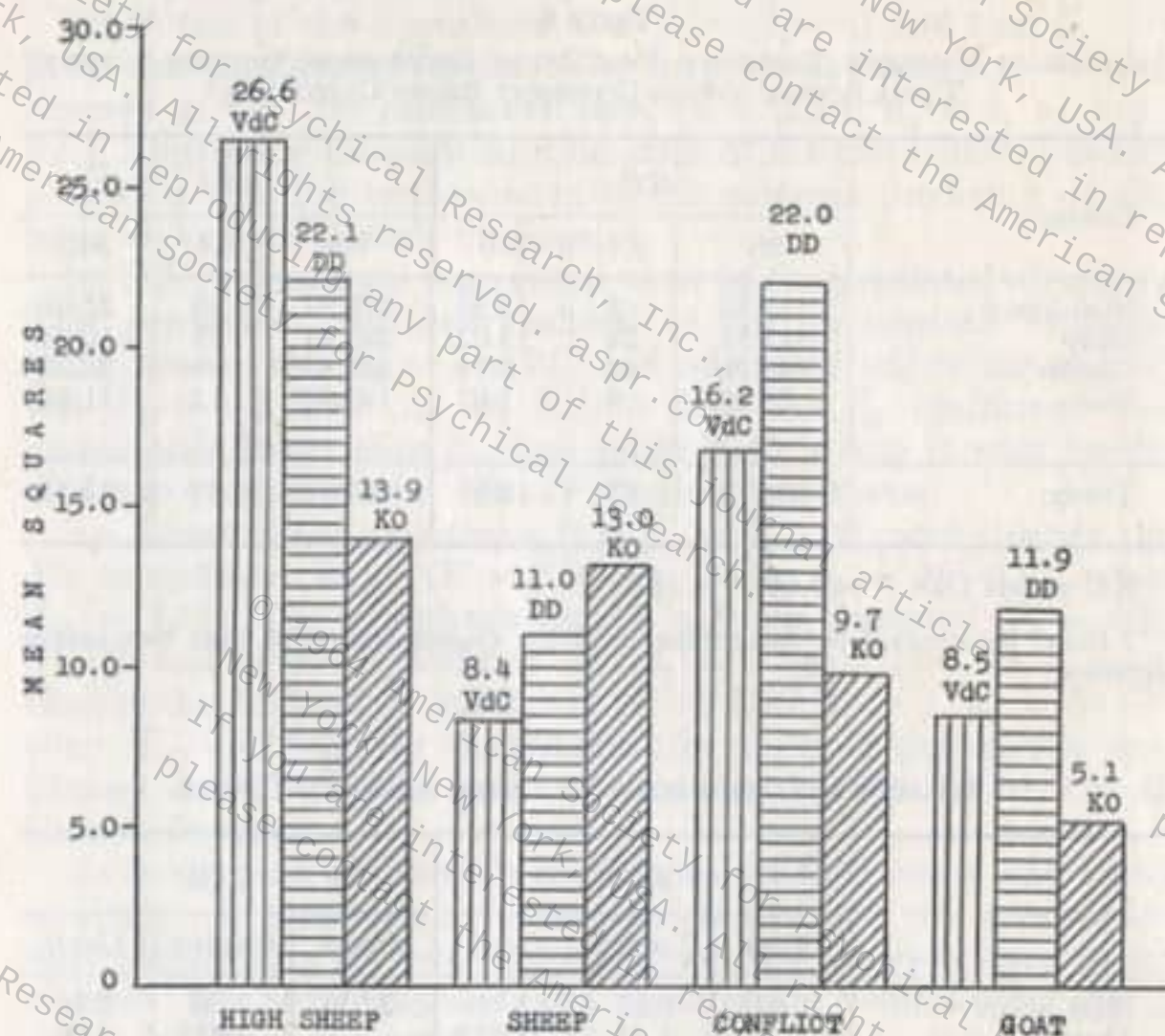


FIG. 5. Between subjects variance in Van de Castle's experiment and in the present series, belief determined by incomplete sentence questionnaire, at least two judges agreeing. Van de Castle's data adjusted by the factor of 2.7.

In this Figure some similarities in between subjects variance patterns can be seen in Van de Castle's and D.D.'s data. The between subjects variance in K.O.'s data does not follow Van de Castle's pattern at all.

Following Van de Castle's suggestion, we also computed between subjects variance in the pooled data of both experimenters. The differences between high sheep *versus* sheep, and high sheep *versus* goats were significant at the .05 level. However, the real probability values might be quite different because the homogeneity of variance requirement of the statistical test used was grossly violated.

Originally it had not been planned to evaluate the between subjects variances according to self-rating scales A and B, but sub-

sequently we decided to compute it for scale A. The results are shown in Tables 7 and 7A. Three out of ten F ratios in K.O.'s data were significant, but two of them represented different patterns from those found by Van de Castle with his self-rating technique. None of D.D.'s F ratios reached significance.

TABLE 7
ANALYSIS OF VARIANCE TABLE FOR EVALUATING DIFFERENCES BETWEEN SUBJECTS' TOTAL SCORES WITHIN DIFFERENT BELIEF CATEGORIES*

Category	KO			DD		
	SS	d.f.	MS	SS	d.f.	MS
High Sheep	222.56	23	9.68	382.98	26	14.73
Sheep	548.48	41	13.38	582.80	36	16.19
Conflict	70.00	7	10.00	349.75	15	23.32
Goats	20.80	4	5.20	44.75	3	14.92
Extreme Goats	9.00	3	3.00	53.50	5	10.70
Totals	870.84	78	11.16	1413.78	85	16.63

KO versus DD: $F = 1.49, P = .05$

* Belief determined by self-rating scale A.

TABLE 7A
F RATIOS BETWEEN BELIEF CATEGORIES SHOWN IN TABLE 7

Category	KO				DD			
	Sheep	Conflict	Goats	Extreme Goats	Sheep	Conflict	Goats	Extreme Goats
High Sheep	1.38	4.03	1.86	3.23*	1.10	1.58	1.01	1.38
Sheep		1.34	2.57*	4.46**		1.44	1.09	1.51
Conflict			1.92	3.33			1.56	2.18
Goats				1.73				1.39

* $P = .05$

** $P = .01$

The correlation between the three questionnaires which were intended to measure belief in ESP was also of interest. Therefore a correlation was made of the incomplete sentence questionnaire results with those of self-rating scales A and B. High positive correlations were found with the following coefficients: $r = .72$, 176 d.f., and $r = .75$, 176 d.f.

The ESP scores of K.O.'s subjects had a low but significant positive correlation with self-rating scale A scores ($r = .30$, $P = .01$). D.D.'s scores had a very slight positive correlation which was not statistically significant ($r = .05$). Other correlations were not performed on the data.

DISCUSSION

The present experiment involved the usual precautions against ordinary perception of the stimulus material as worked out by Schmeidler (5), supplemented by Van de Castle and White (9), and used by Anderson and White, Rilling, and others in school experiments (8). These precautions have withstood cross-fire and are generally considered to be adequate. There were two independent records: two copies of the target order and two copies of the subjects' calls. They were checked twice by two assistants who did not know the belief ratings at the time of checking. A number of subjects had to be eliminated from further evaluations because of incomplete ESP tests or questionnaire responses, but elimination was according to a set of objective rules and carried out without knowledge of scoring level on either ESP tests or belief questionnaires (see Footnote 6).

Although the present experiment was essentially a repetition of the earlier work of Schmeidler and Van de Castle, the procedures were not identical with theirs. The number of runs per subject in our experiment was four as compared with eight in Van de Castle's experiments and an average of nine in Schmeidler's; and attitudes toward ESP were ascertained after the ESP testing rather than before. These differences are mentioned because they might have been important in affecting scoring and variance levels.

We will first consider the ESP scores in relation to the measures of belief in the existence of ESP. From inspection, as well as from statistical analysis, it is apparent that the effects are much more clearly expressed in K.O.'s data than in D.D.'s, although the disbelief (goat) categories might form an exception (Figures 1 and 2). Moreover, the hypothesis brought out in previous work (2, 5) that a decreasing belief in ESP is accompanied by a roughly linear pattern of decline in scores is demonstrated by K.O.'s subjects. The ESP scores of D.D.'s subjects show a slightly curvilinear distribution in relation to the belief variables; however, the trend was not strongly enough expressed to justify a test of curvilinear regression. The main differences between K.O.'s and D.D.'s subjects are in the strong belief (high sheep) category, where K.O.'s subjects scored significantly above and D.D.'s

subjects below mean chance expectation (Figures 1 and 2). Significant differences between K.O.'s and D.D.'s data emerge only in the combined sheep categories, but not in the combined conflict and goat categories when we take each subject as a unit, scoring either above or at or below chance expectation, in a 2×2 table evaluated by chi-square analysis. The same trends prevail regardless of whether the incomplete sentence questionnaire or self-rating scale A is used as the measure of belief. The multiple regression analysis also underscores experimenter differences, i.e., a roughly linear relationship between ESP scores and two belief measures in K.O.'s data and no such relationship in D.D.'s data.

How can we explain these differences between the two experimenters in view of the fact that the talks both gave were based on the same outline and illustrated with the same slides? We can speculate that the conflict and goat subjects responded to the talk in the same fashion for both experimenters, while D.D. irked the sheep and reversed their motivation. An exploratory analysis of the subjects' emotional arousal¹⁰ indicated that K.O.'s combined sheep scored on about the same level regardless of degree of arousal. In contrast, subjects in D.D.'s sheep groups performed in a curvilinear fashion: those moderately aroused scored above mean chance expectation whereas the not aroused and the highly aroused scored below. The goats in both experimenters' groups scored below the chance line regardless of their emotional arousal (Figure 6).

Have we any clues as to why these different reactions should have occurred? A few impressions can be put on record. The assistant (S.V.S.) who helped in the experiments and heard both lecturers remarked that whereas K.O.'s speech was presented in a relaxed, humorous, and confident manner, D.D. was rather tense, less secure about the subject matter, and very seriously intent on getting the audience to accept his point of view. In self-observations by both experimenters it was discovered that K.O. loves to give talks and thoroughly enjoys such occasions; D.D., on the other hand, did not enjoy giving talks at that time and became nervous when confronting an audience. Both experimenters are definitely convinced of the existence of ESP. If there is any difference in

¹⁰ As a post-mortem exploratory measure K.O., using a blind procedure, evaluated all subjects as to their emotional arousal on the basis of their incomplete sentence questionnaire responses. Emotionally-charged words and statements were tabulated according to the frequency and intensity of the emotion expressed. Sharp contradictions in statements were also counted. The subjects' emotional arousal was judged on a three-point scale. In order to segregate the emotionally aroused subjects with sharp distinction the benefit of the doubt was given to the non-aroused class (arousal 1), and as a result more than half (93 subjects) were placed in this category.

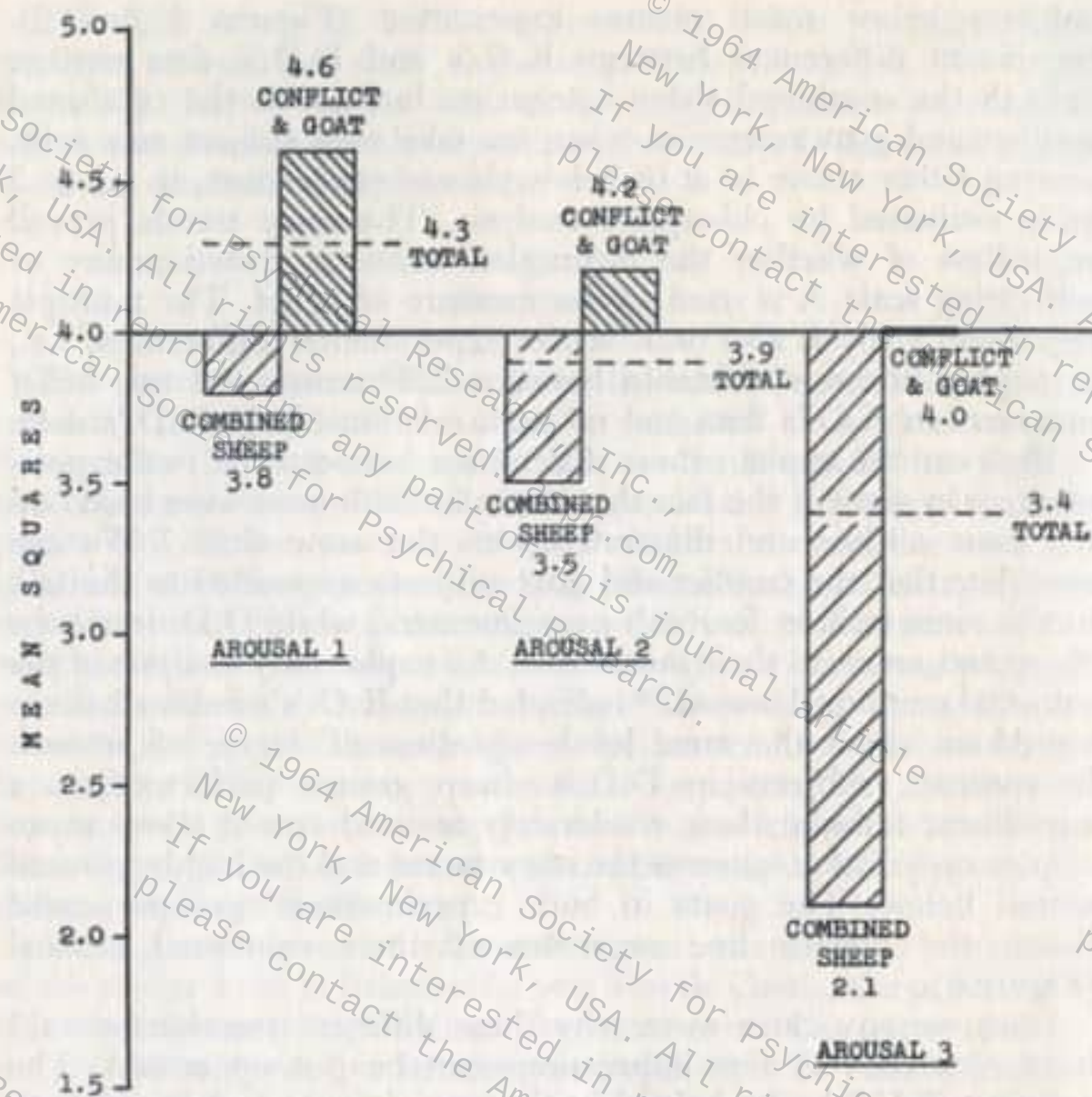


FIG. 6. Within subject variance of the pooled data of both experimenters according to levels of emotional arousal.

level of belief between the two, D.D.'s belief would be more intensely positive than K.O.'s. Is it possible that the experimenter who enjoys the lecture session for the fun of it gets good scores, while the one for whom it is a difficult task fails to get positive scores? This could be one variable to consider in further explorations. If we could finally specify the relevant factors in experimenters which make for successful ESP testing, a major skirmish in the battle for repeatable experiments would be won. Personality and motivational factors would seem to need consideration in experimenters as well as in subjects.

Another major concern of the present experiment was to test Van de Castle's hypothesis that within subject variance would be lower in the combined sheep than in the pooled conflict and goat subjects. With the F test that Van de Castle used, our results

were confirmatory at the .05 level of significance. However, this finding must be taken with caution since when we used Van de Castle's chi-square measure, taking each subject as a unit, we obtained chance results. The 2×2 table of subjects in both pooled belief categories having variance either above or below the mean was the basis of the chi-square test. The results of the chi-square analyses were not significant when using both belief measures (incomplete sentence questionnaire and self-rating scale A). The reversal of the between subjects variance trend (sheep, high; conflict-goat, low) in K.O.'s data in relation to self-rating scale A was completely unexpected. If this reversal is a real effect and not just chance coincidence, it would seem to accentuate differences between the two belief measures. The incomplete sentence questionnaire and self-rating scale A are highly correlated ($r = .72$, 176 d.f.); therefore, they seem to measure some belief factors similarly. But it is possible that one of them might be a sensitive measure of some factors which the other measures poorly. In view of our results, the two instruments might have registered differently the dimensions responsible for stabilizing ESP (within subject variance), while registering similarly those dimensions which affect scoring level.

A multiple regression analysis was performed on the within subject variance. This analysis clearly indicated the difference between the two belief measures as well as the differences between the data of the two experimenters. The multiple regressions approached significance only in D.D.'s data, where the regression coefficients of both belief measures had similar signs. Generally speaking, D.D.'s variance results were more in conformity with Van de Castle's findings than were K.O.'s.

We also compared the within subject variance of Van de Castle and the two present experimenters (Figure 4). The widest divergence is apparent in the goat category; the most uniform trend appears in the conflict category. In Van de Castle's own experiment (7), the goats' within subject variance was also very unstable, varying from series to series from 12.88 to 40.69. His conflict group was remarkably consistent throughout his seven series, with the variance ranging only from 30.04 to 38.99. We might speculate that the motivational conflict in goats is entirely dependent on the test situation. Van de Castle entertained the idea that the conflict of the disbeliever is between a firm conviction that ESP tests are useless and a waste of time on the one hand, and a social obligation to cooperate with the experimenter on the other. If this is the case, the social obligation would be felt quite differently in relation to different experimenters, and the conflict, therefore, would depend largely upon the experimenter's personality and approach.

The conflict group subjects seem to have an inner conflict between their own doubts and beliefs; that is, conflicting value systems. If this is so, the conflict would be relatively stable and less dependent on the immediate, external situation.

Figure 6 illustrates the emotional arousal factor in relation to within subject variance in the different belief categories (data of both experimenters pooled). It was a surprise to find that the variance was negatively related to the arousal intensity; that is, more arousal, less variance. This trend is distinctly expressed in the combined sheep category, and to a lesser degree also in the pooled conflict and goat category. The high sheep and goat categories taken singly do not show a linear decline; this may be due to the very small number of subjects with extreme beliefs split among three arousal categories.

The emotional arousal dimension was introduced as a post-mortem analysis and therefore was not evaluated statistically. But if this trend is a real effect, we could speculate that emotional arousal, to the extent that it is expressed in verbal responses, might stabilize the ESP scoring level, at least in the combined sheep categories. In spite of the fact that K.O.'s sheep scored above and D.D.'s below mean chance expectation, the between subjects variance decreased quite similarly in both experimenters' data. It would appear that emotional arousal stabilized both psi hitting and psi missing.

Between subjects variance in Van de Castle's data indicates a definite pattern: very high scoring variation in the high sheep category, moderately high in the conflict group, and very low in the sheep and goat categories. In other words, strong believers did not respond in a uniform manner to Van de Castle as an experimenter and had scores scattered from very high to very low. In contrast, moderate believers and disbelievers responded as a homogeneous group and had little variation in total scores (Figure 5).

D.D.'s between subjects variance followed the same pattern as Van de Castle's, with the magnitude of the variance being in the same rank order, but of the six comparisons only one reached the .05 level of significance (Table 6A). K.O.'s between subjects variance showed no resemblance to either D.D.'s or Van de Castle's patterns.

Because of the low F values obtained, repetition of Van de Castle's findings in relation to between subjects variance is not claimed. If the trends in our data represent something more than chance coincidence, we might speculate that, like D.D., Van de Castle annoyed some of the high sheep and accentuated the motivational conflicts of the goats. This hypothesis is supported by

another unpredicted finding in the data of the present experiment: D.D.'s between subjects variance is significantly higher than K.O.'s. It would be interesting to survey past ESP experiments to see whether low between subjects variance is consistently associated with significant above-chance scores and high between subjects variance (where psi-hitting might be cancelled out by psi-missing) is an adjunct of non-significant scores.

The scores of the incomplete sentence questionnaire correlated highly with those on both the A and B self-rating scales. However, only self-rating scale A and the incomplete sentence questionnaire seemed useful in separating psi-hitting from psi-missing subjects. Multiple regressions on K.O.'s data indicated that these two measures used jointly were more effective than either one alone. Self-rating scale B might not be so much a measure of the subject's belief in personal ESP ability as such as a measure of the interaction between such belief and his manner of self-description, i.e., boastfulness, overcautious modesty, and conformity. The incomplete sentence questionnaire was more effective than self-rating scale A in relation to between subjects variance.

The findings of this experiment indicate the great complexity of the motivational factors in ESP. The decisive importance of the subject's attitude toward the experimenter and the test situation was revealed in a comparison of K.O.'s and D.D.'s results. However, no direct measure of such attitudes was obtained. To explore this factor modern tools of psychological scaling, e.g., Osgood's semantic differential method (3), could be applied to subjects and experimenters in future research. Complex research designs also require complex multivariate evaluation methods, such as have been introduced here in a rather modest way.

REFERENCES

1. Adcock, C. J., and Quartermain, D. "Some Problems in Group Testing of ESP." *Journal of Parapsychology*, Vol. 23, December, 1959, 251-256.
2. Mangan, G. L. *A Review of Published Research on the Relationship of some Personality Variables to ESP Scoring Level*. Parapsychological Monograph No. 1. New York: Parapsychology Foundation, Inc., 1958.
3. Osgood, C. E., et al. *The Measurement of Meaning*. Urbana: University of Illinois Press, 1957.
4. Schmeidler, G. R. *ESP in Relation to Rorschach Test Evaluation*. Parapsychological Monograph No. 2. New York: Parapsychology Foundation, Inc., 1960.

5. Schmeidler, G. R., and McConnell, R. A. *ESP and Personality Patterns*. New Haven: Yale University Press, 1958.
6. Solley, C. M., and Murphy, G. *Development of the Perceptual World*. New York: Basic Books, Inc., 1960.
7. Van de Castle, R. L. "Differential Patterns of ESP Scoring as a Function of Differential Attitudes toward ESP." *Journal A.S.P.R.*, Vol. 51, April, 1957, 43-61.
8. ———. "A Review of ESP Tests Carried Out in the Classroom." *International Journal of Parapsychology*, Vol. 1, Spring, 1959, 84-102.
9. Van de Castle, R. L., and White, R. R. "A Report on a Sentence Completion Form of Sheep-Goat Attitude Scale." *Journal of Parapsychology*, Vol. 19, September, 1955, 171-179.

A.S.P.R.
880 Fifth Avenue
New York 21, N. Y.

APPENDIX

Incomplete Sentence Questionnaire

1. I believe that hunches often
2. Stories we hear about telepathy
3. People who are skeptical of ESP are
4. Personally clairvoyance seems
5. The scientific evidence for parapsychology
6. I think telepathy
7. I doubt that ESP
8. I do not believe that clairvoyance

9. A careful study of parapsychology would show
10. It is impossible for ESP

Self-Rating Scales A and B

We would like to know your opinions on the following questions. Please check that answer which best applies to you. Be as accurate and reliable as possible.

- A. Do you think some people are sometimes able to obtain information by ESP?
Sure about it —, this is a possibility —, it might be so or not —, it seems unlikely —, it is impossible —.
- B. Do you believe that you are sometimes able to get impressions by means of ESP?
Sure about it —, this is a possibility —, it might be so or not —, it seems unlikely —, it is impossible —.