PROBABILITY ESTIMATION IN A CHANCE TASK WITH CHANGING PROBABILITIES

A Thesis
Presented to
the Faculty of the Department of Psychology
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Science

By
Alice L. Bane
August, 1973

ACKNOWLEDGMENT

The suggestions and criticisms of Eric R. Brown were of great value in the conduct of this study, and are sincerely appreciated.

PROBABILITY ESTIMATION IN A CHANCE TASK WITH CHANGING PROBABILITIES

An Abstract of a Thesis
Presented to
the Faculty of the Department of Psychology
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Science

By
Alice L. Bane
August, 1973

ABSTRACT

Subjects made sequential estimates of the probability of occurrence of one event in a chance-controlled binary-event sampling situation. Subjects whose samples were drawn from populations in which the probability of occurrence of the event in question increased tended to overestimate probabilities, while subjects whose samples were drawn from stable probability populations tended to underestimate probabilities. A second phase in which probabilities were stable for all subjects resulted in overestimation by subjects who had been in the changing probability conditions and underestimation by subjects who had been in the stable condition, although the different probability levels reached in the different conditions may have confounded these results to some extent. The results of this study show that overconfidence can be obtained in a chance controlled task in the presence of increasing probabilities, and indicate that the overconfidence of success in skill situations reported in previous studies may be due in part to the rising probabilities typically associated with skill tasks. Some suggestions for future investigations of subjective probabilities in the presence of changing objective probabilities are presented.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Previous research in the area of internally controlled versus externally controlled tasks, i.e., skill versus chance tasks, indicates that people tend to be overconfident of uncertain outcomes which depend upon their own performances. In those instances in which skill tasks were involved, a strong tendency for subjective probabilities to exceed objective probabilities has been reported (Howell, 1967, 1971, 1972). However, the literature concerning external uncertainty reveals that in the instances where external tasks were involved, subjective probability (SP) did not regularly exceed objective probability (OP) (Peterson and Beach, 1967).

In two of Howell's studies (1967, 1972) subjects were required to produce indirect measures of subjective probability concerning skill tasks by making choices between skill and chance tasks when the objective probabilities for the chance tasks were known. The chance task involved the use of spinners with pie segments colored to represent proportions from .10 to .90 in steps of .10. The OP associated with a spinner was clearly marked beneath that spinner. In the 1972 study, each skill task (dart throwing for various rings) was pitted against a series of spinner values from .10 to .90, and the subjective indifference point, expressed in "objective" spinner values, was taken as an index of SP. These inferred or indirectly measured SP estimates were compared to direct measures of SP (subjects estimated probabilities) and to OP. OP was defined by Howell as asymptotic accuracy. Indirectly measured SP exceeded directly measured SP, which in turn exceeded OP for the skill tasks studied.

To summarize, subjects appear to be overconfident of success in a task in which the outcome depends upon one's own performance, but this overconfidence is not regularly found when the outcome depends upon chance. This brings up the problem of the difference between skill and chance events. What are the characteristics of a skill event, characteristics that are not found in a chance event, that result in this overconfidence? One likely characteristic, noted by Howell (1971) and others (e.g., Feather, 1959), is the possible presence of a utility factor. It may be that when skill tasks are compared with chance tasks there is a preference for skill tasks simply because these tasks have more utility associated with them. Feather (1959) hypothesizes that the attainment attractiveness, which he states is analogous to utility, should vary inversely with the associated success probability, and that this assumed covariation should tend to be more apparent in ego-related than in chance-related situations.

A related characteristic of skill tasks is that they are typically associated with improvement. That is, when a person is first introduced to a skill task, he typically expects to do better on later trials than on earlier trials—he expects improvement, or rising probability of success. By contrast, chance tasks, at least those chance tasks used in laboratory experimentation, typically do not carry with them the promise of improvement. In particular, the chance tasks used by Howell were constant at each probability level used. Therefore, in choosing between chance and skill tasks, Howell's subjects were choosing between a task which had been and would remain constant at a definite level of probability (the chance task) and one in which throughout its history in the experiment and in generalization from other skill tasks there had been improvement. This portential for improvement may account for some of the overconfidence found in connection with skill tasks. Although Howell (1971, 1972) states that his subjects had entered the asymptotic stage, subjects may not have been aware or may not have believed that the potential for improvement was no longer present. It is possible that some lingering effect of the improvement which had taken place during the earlier phase of the learning process contributed to overconfidence during the asymptotic phase.

A chance controlled binary-event task in which the probability of one event incrases while the probability of the other event decreases during the course of the experiment seems to approximate a skill task in that improvement in one event is part of the history of the task. Restle and Greeno (1970) present various learning models, and refer to binary-event urn sampling situations as illustrations of the models. If a draw of one color is regarded as a correct response and a draw of the other color is regarded as an incorrect response, then "learning", or a rising probability of correct response draws, can occur through any one of several sampling schemes, each of which results in an increase in the relative proportion of the correct response color. The parameters of these models can be varied to fit data from actual learning experiments, which demonstrates that a chance task can be used to simulate the improvement typically found in a learning situation. The use of a binary-event task of this type in a situation in which subjects estimate probabilities might result in SP for the increasing probability event which exceeds OP, an occurrence which has not regularly been associated with chance tasks. If this were to occur, it would indicate that part of the overconfidence associated with skill tasks arises not from uncertainty whose locus is internal rather than external, but from the presence of a change or improvement factor, which is typically associated with internal but not with external tasks.[1] Or, to approach the matter from another angle, it could be said that this change factor describes one aspect of internal tasks not found in the external tasks studied, and is part of the difference between them. The purpose of this experiment was to test the hypothesis that in a binary-event chance task with changing probabilities, SP for the event whose probability increases over trials will exceed OP for that event during an increasing probability phase and during a following asymptotic phase.

II. METHOD

72 undergraduate volunteers from introductory psychology courses at the University of Houston were used as subjects, and were randomly assigned in order of appearance to 4 groups of 18 subjects each. Although no attempt was made to balance sex of subjects in groups, random assignment resulted in no major inbalance in any group. The task consisted of drawing red and white poker chips from a cloth bag and estimating probabilities of drawing a specified color chip on selected trials.

The individual testing procedure, which required approximately one-half hour per subject, took place in a small interviewing room. The room was furnished with a table, two chairs on opposite sides of the table, a low stand below the table level beside the experimenter's chair, and shelves for supplies (extra pencils, materials) behind the experimenter. The subject's books and other belongings were placed on the shelves, and the subject was seated in the subject's chair. The experimenter read the instructions aloud, and offered to re-read any parts the subject wanted to hear again. (Few subjects requested additional readings.) The subject was then asked to reach into the bag and to draw one poker chip at a time. The correct replacement or removal, to be discussed below, was carried out immediately, and the chips in the bag were mixed before the next draw.

The subject was presented with a bag containing a poker chip population in which the initial frequencies of the two colors were 35 chips of one color and 15 chips of the other color. Colors were counterbalanced so that half the subjects in each group were presented populations in which red chips were more frequent, and half were presented populations in which white chips were more frequent. The subject was asked to draw one-chip samples from the population, and, at 5 sample intervals, to make estimates concerning the probability of drawing a particular colored chip on the next draw. The subject made his estimate by making a pencil mark on a horizontal line 150 mm. in length where the left end point corresponded to a probability of zero and the right end point corresponded to a probability of 1.00. The distance from the left end point of the line to the point of intersection of the subject's mark with the line was later measured to the nearest millimeter, and this distance was divided by 150 mm. (the entire length of the line) to provide a probability value. Peterson and DuCharme (1967) used a similar scale-and-pointer method of indicating probability estimates. Because of subjects' greater familiarity with percentages than with probabilities, the line was actually labeled 0% at the left end and 100% at the right end. No other orientation points were provided.

Sampling schemes differed for the 4 groups. For Group C (Constant), all samples were replaced, so that the relative frequencies, hence the probabilities, of each color remained constant. For Group A (Add), each chip drawn of the color which was more frequent in the original population (the color whose frequency was 35) was replaced, while each chip drawn of the color which was less frequent was replaced along with another chip of that color. For Group S (Subtract), each chip drawn of the originally more frequent color was removed, while each chip drawn of the originally less frequent color was replaced. For Group A-S (Add-Subtract), each chip drawn of the originally more frequent color was removed, while each chip drawn of the originally less frequent color was replaced along with another chip of that color. Therefore, the probabilities in Group C remained at constant levels of .7 and .3, while in Groups A, S, and A-S the probability of drawing the originally less frequent color increased while the probability of drawing the originally more frequent color decreased. The probability of drawing the originally less frequent color increased from .3 to an average of .582 for Group A, to .858 for Group S, and to .895 for Group A-S. The change in probability was least for Group A and greatest for Group A-S. Increased probabilities of drawing the originally less frequent color resulted from draws of that color in Group A, from draws of the other color in Group S, and from any draw in Group A-S.

After making 5 draws, the subject was asked to estimate the probability of drawing a chip of a specified color on the next draw. The specified color was that of the lower original frequency, the probability of which increased during the experiment, and which will hereafter be referred to as the "focal event". Estimations were taken at 5 draw intervals throughout the experiment. After the 5th, 10th, 15th.....nth draws, the subject estimated the probability of drawing the focal event on the 6th, 11th, 16th.....n+1 draws, respectively. Estimates were set aside as they were made; subjects were not permitted to refer to previous estimates. Probabilities continued to change while the subject made 75 draws and 15 estimates, after which the changing probability groups were shifted to a no-change phase, which was included in order to approximate the asymptote of a learning curve. The shift to the no-change phase was accomplished by abandoning the sampling scheme that resulted in changing probabilities (Phase I) and replacing it with a simple sampling with replacement scheme (Phase II). The subject was informed of the shift. The experimenter recorded draws for each subject, recording Phase I on the first side and Phase II on the second side of the record sheet. Turning the sheet provided a pause at the point of shift from Phase I to Phase II for all groups, including Group C, for which there was no actual shift. Phase II continued for six more estimations (30 draws) making a total of 21 estimations (105 draws) in the entire procedure.

The subject received information about the sampling scheme used in his particular condition (all chips will be replaced, red chips will not be replaced, etc.), but received no information concerning initial proportions or size of the population. All replacements and/or removals were carried out in the presence of the subject, but no special efforts were made to draw his attention to these procedures. Chips which had been removed or which were to be added were kept on a low stand beside the experimenter where they were not in the line of sight of the subject.

Each subject's draws determined the actual sequence of events for that subject. The between-subjects variability inherent in such a procedure was thought to be more acceptable for the present purposes than the possible confounds attending a predetermined sequence of events. It was reasoned that the use of predetermined sequences might have led to the suspicion of the part of subjects that the task was controlled by the experimenter, that patterns might be found, or that some skill factor might be involved.

III. RESULTS AND DISCUSSION

Comparison of SP and OP

In Figures 1—4, the average OP (the actual proportion of focal event chips in the bag at the time of estimation averaged across the 18 subjects in each group), the average SP, and the average proportion of focal event chips drawn by subjects in each group over the five draws immediately preceeding the estimation are illustrated by blocks of three estimation trials. The OP curves for Groups A, S, and A-S, the three changing groups, resemble learning curves, and could be said to approximate the learning of skill tasks at different rates. SP curves lie above OP curves for each of the groups in which OP increased, while SP is below OP for Group C (Figure 1).

Average overconfidence (SP-OP) for the different groups is illustrated by blocks of three estimation trials in Figure 5. There appears to be overconfidence in the three changing probability groups (A, S, and A-S), and underconfidence in the one constant probability group (C). The average SP-OP difference was computed for each group for Phase I and Phase II separately, and was tested using Student's t-Ratio (two-tailed test). (The Student's t-Ratio results are reported in Table 1.) All SP-OP averages were found to be significantly different from zero, with average SP-OP for Group C negative in both phases, and average SP-OP for Groups A, S, and A-S positive in both phases. These findings support the hypothesis that the overconfidence usually associated with skill tasks may be due in part to the presence of changing probabilities.

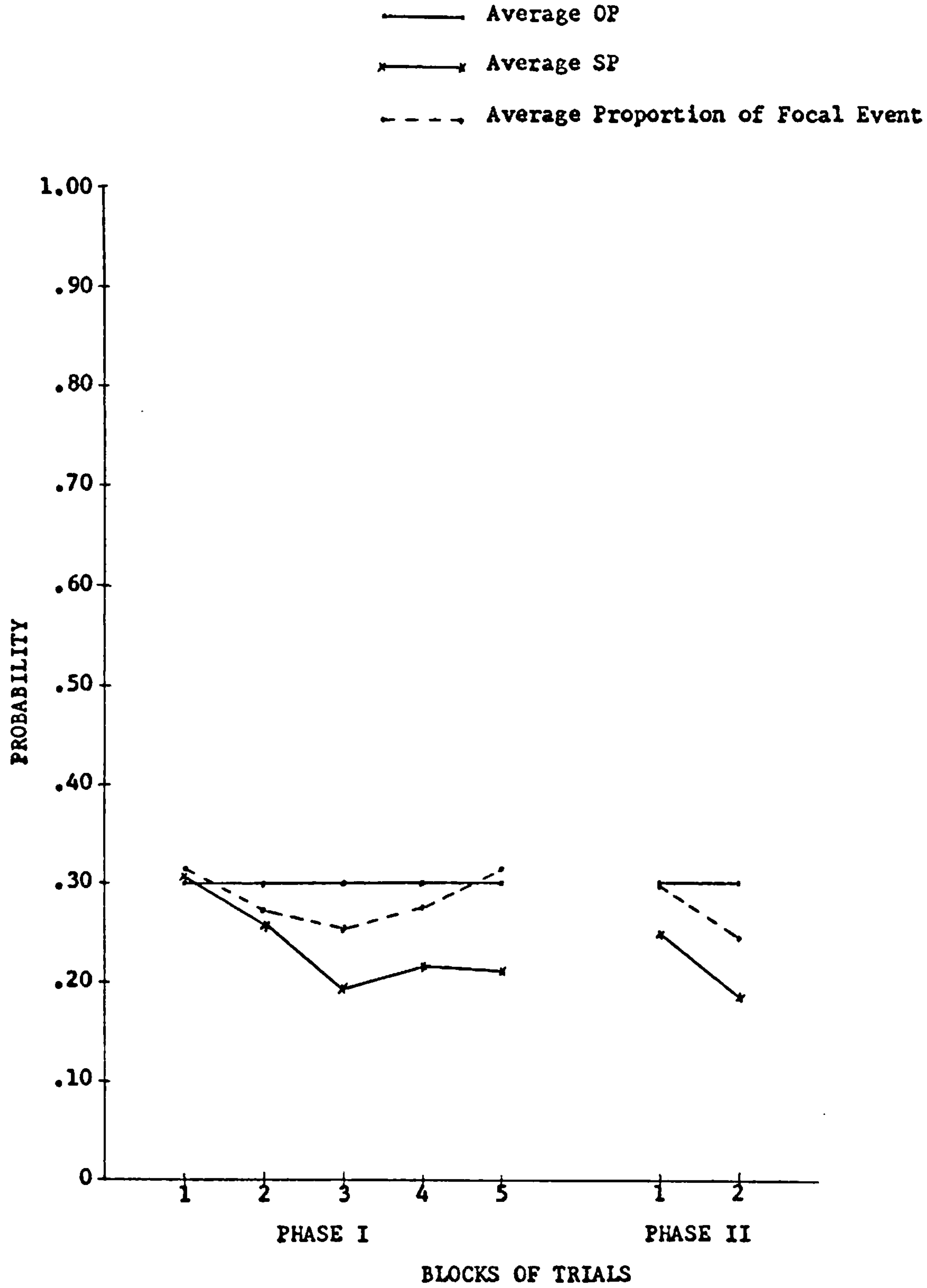

Figure 1. Average OP, Average SP, and Average Proportion of Focal Event Chips drawn in the five draws immediately preceeding an estimation, by blocks of three estimation trials; Group C.

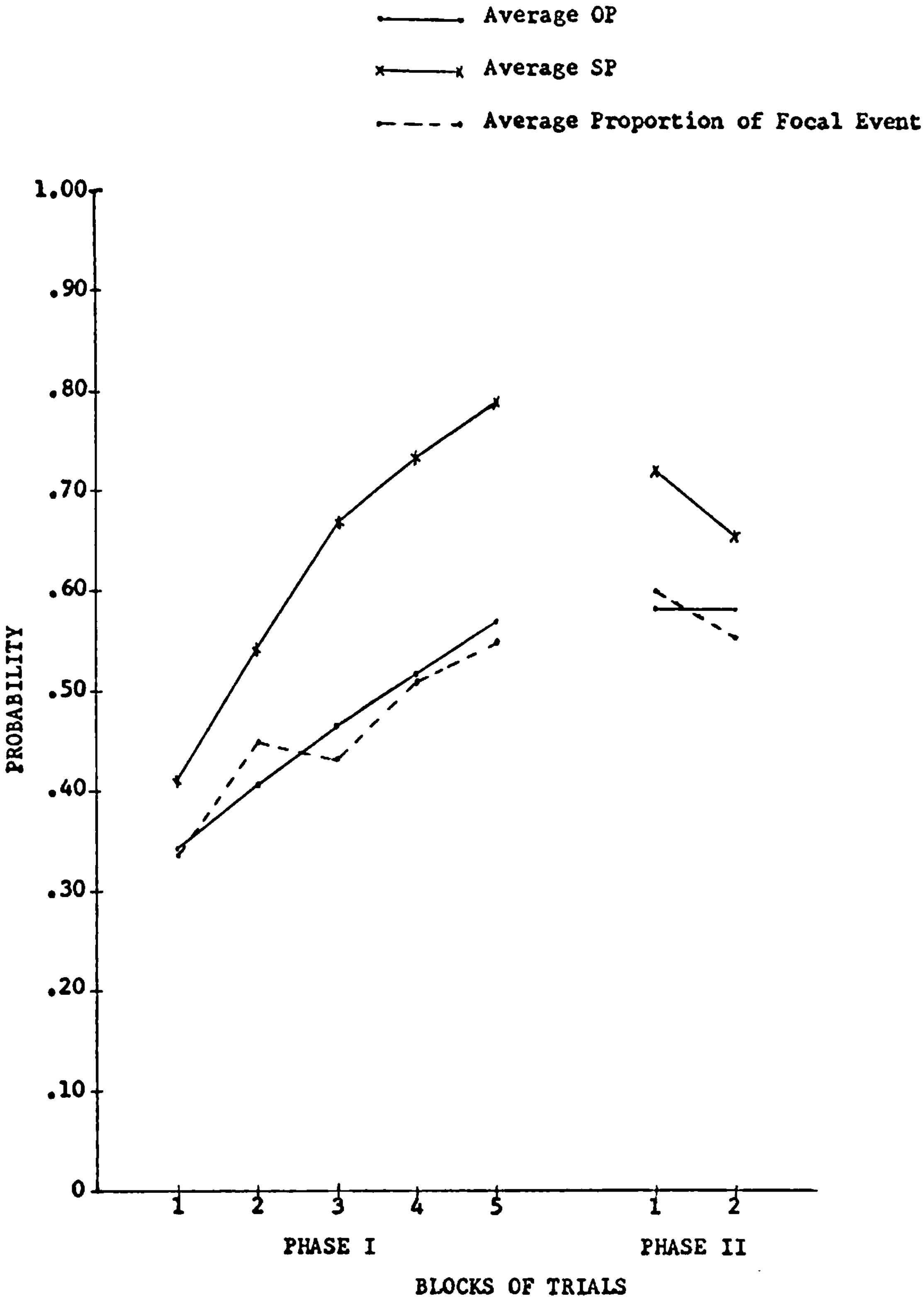

Figure 2. Average OP, Average SP, and Average Proportion of Focal Event Chips drawn in the five draws immediately preceeding an estimation, by blocks of three estimation trials; Group A.

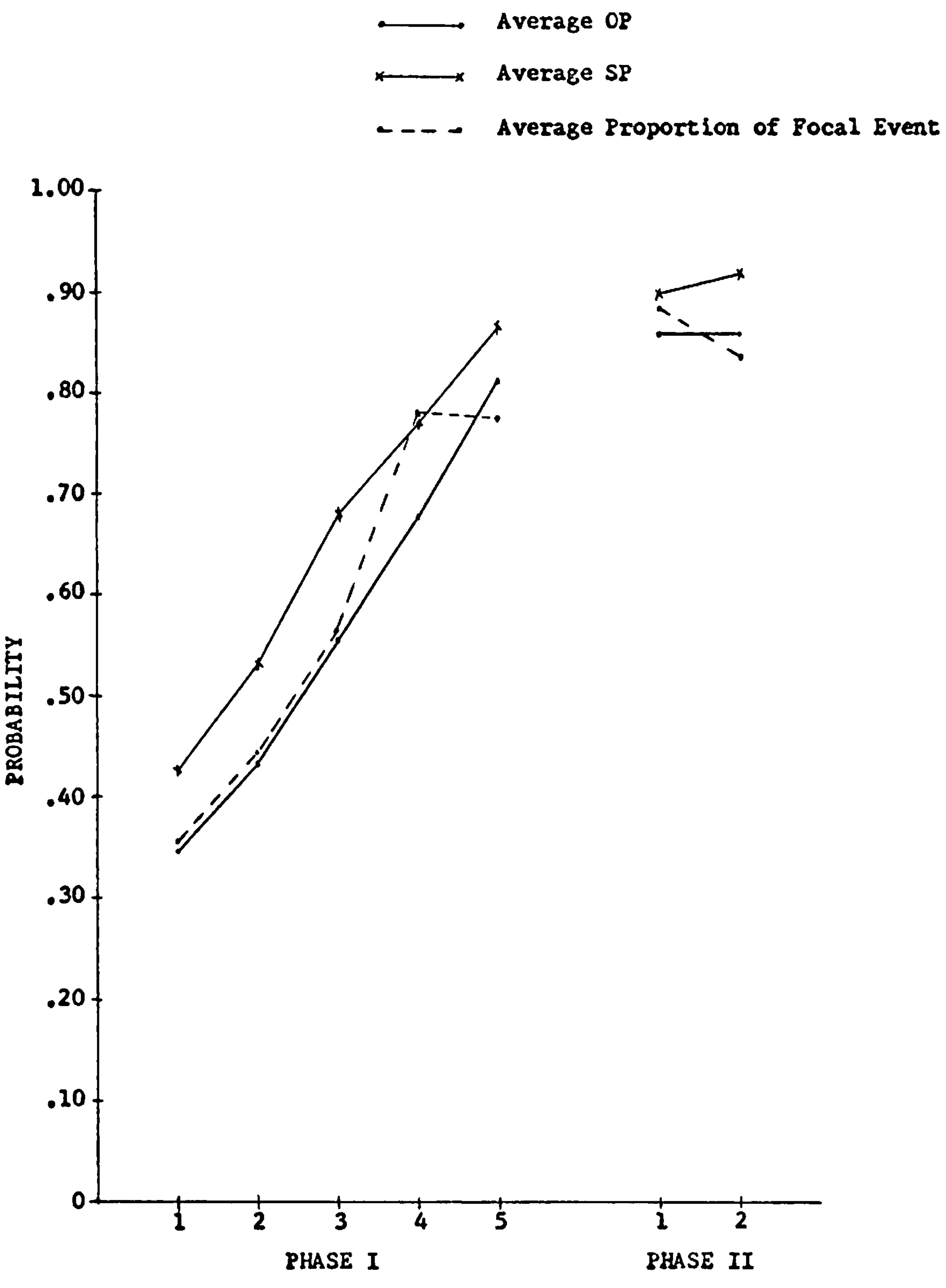

Figure 3. Average OP, Average SP, and Average Proportion of Focal Event Chips drawn in the five draws immediately preceeding an estimation, by blocks of three estimation trials; Group S.

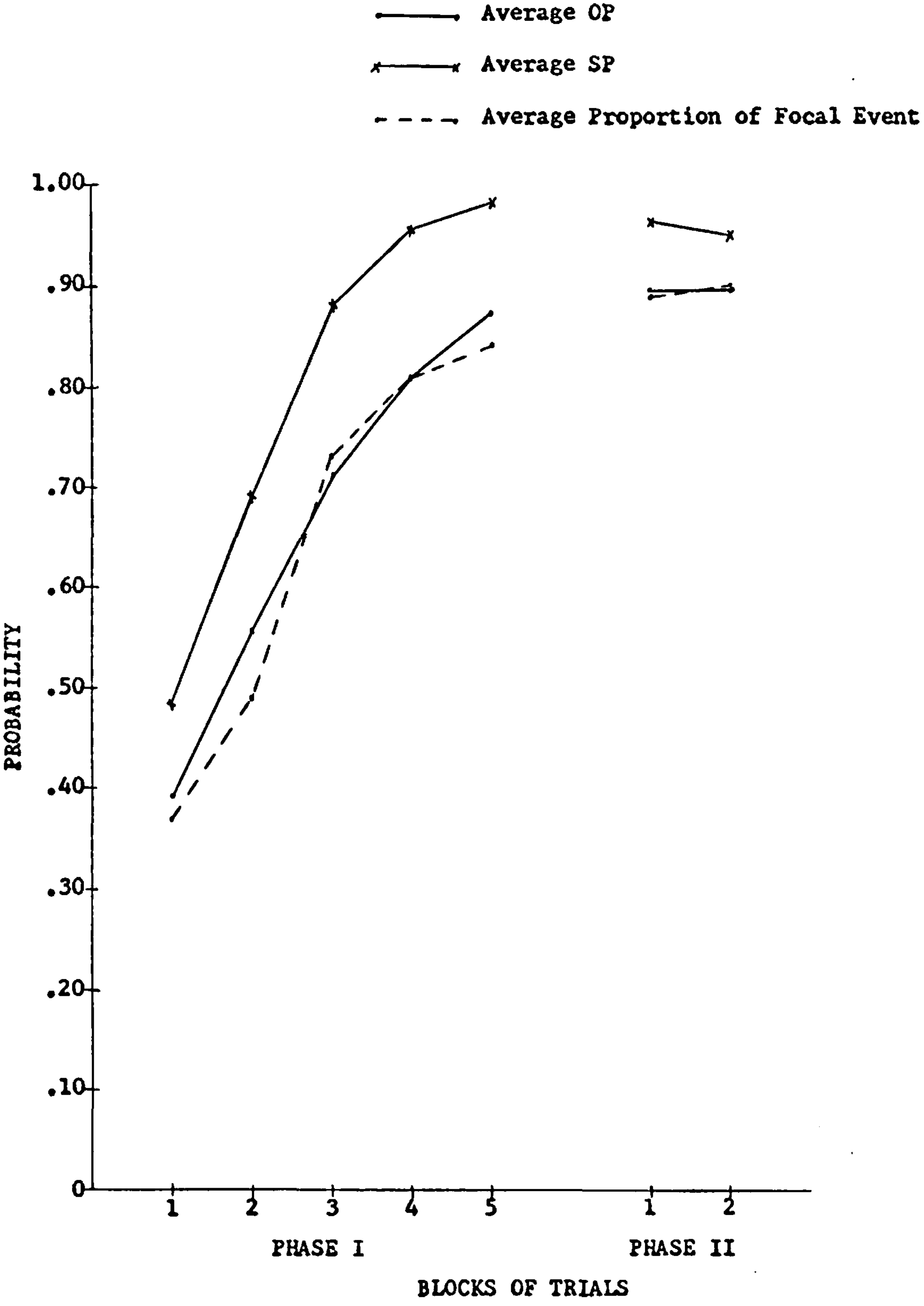

Figure 4. Average OP, Average SP, and Average Proportion of Focal Event Chips drawn in the five draws immediately preceeding an estimation, by blocks of three estimation trials; Group A-S.

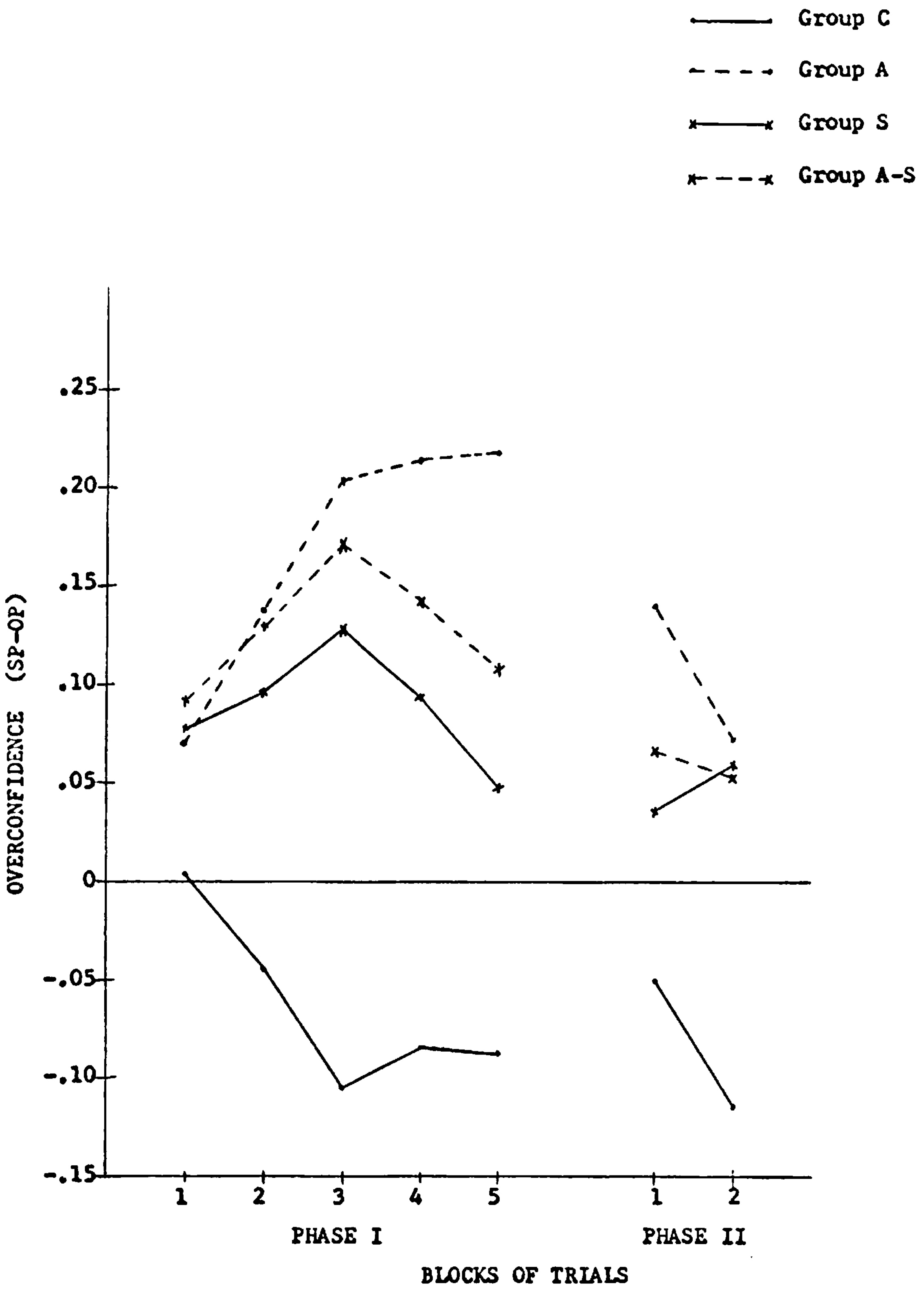

Figure 5. Average Overconfidence (SP-OP) by blocks of three estimation trials.

Table 1. Results of Student's t-Ratio Testing the Null Hypothesis $\mu_{\overline{SP-OP}} = 0$.

PHASE I

	Group C	Group A	Group S	Group A-S
t (df 17)	-2.666	5.124	2.678	5.215
probability (two tailed)	$p < .02$	$p < .001$	$p < .02$	$p < .001$

PHASE II

	Group C	Group A	Group S	Group A-S
t (df 17)	-2.670	2.838	3.613	2.540
probability (two-tailed)	$p < .02$	$p < .02$	$p < .01$	$p < .025$

Comparison of SP-OP for Groups A, S, and A-S During Each Phase

The appropriate analysis of variance for a two-factor experiment (Groups X Blocks) with repeated measures on one factor (Blocks) (Winer, 1962) was performed to determine whether the three changing probability groups differed in amount of overconfidence. The Groups X Blocks ANOVA on Phase I resulted in a blocks main effect ($F_{4, 204} = 3.277$, $p < .05$), but no significant group main effect or interaction. (See Table 2 for a summary of the ANOVA on Phase I.) The blocks main effect reflects the tendency illustrated in Figure 5 for SP-OP for all three changing probability groups to increase over the first three blocks in Phase I, after which SP-OP for Group A tends to level off while SP-OP for Groups S and A-S tends to decline.

The Groups X Blocks ANOVA on Phase II resulted in no significant F ratios, although the groups X blocks interaction approaches significance at the .05 probability level. Average SP-OP for appears to decline during Phase II, while average SP-OP for Groups S and A-S does not appear to change markedly. (See Table 3 for a summary of the ANOVA on Phase II.)

Individual Differences

Large individual differences were encountered in the present study, and were alluded to by Howell (1972) in connection with SP-OP for the skill task. SP-OP group means and standard deviations were computed for each of the three changing probability groups, and it was found that in both Phase I and Phase II a distance of one standard deviation from the mean of any one group includes the means of the other two groups. (Group means and standard deviations are presented in Table 4.) The large

Table 2. Groups (A, S, and A-S) by Blocks Analysis of Variance, Phase I.

SOURCE OF VARIATION	SS	df	MS	F	
Between Subjects	4.5588	53			
Groups	.2857	2	.1428	1.704	ns.
_Ss_ Within Groups	4.2731	51	.0838		
Within Subjects	4.2617	216			
Blocks	.2463	4	.0616	3.277	*
Groups X Blocks	.1844	8	.0230	1.229	ns.
Blocks X _Ss_ Within Groups	3.8310	204	.0188		

* $p < .05$

Table 3. Groups (A, S, and A-S) by Blocks Analysis of Variance, Phase II.

SOURCE OF VARIATION	SS	df	MS	F
Between Subjects	1.788110	53		
Groups	.065249	2	.032624	.966 ns.
_Ss_ Within Groups	1.722861	51	.033782	
Within Subjects	.359951	54	.006666	
Blocks	.009167	1	.009167	1.495 ns.
Groups X Blocks	.037978	2	.018989	3.010 ns.
Blocks X _Ss_ Within Groups	.312806	51	.006133	

Table 4. SP-OP Group Means and Standard Deviations

	PHASE I		PHASE II	
	Group Mean	Standard Deviation	Group Mean	Standard Deviation
Group A	.169	.136	.106	.154
Group S	.089	.137	.085	.097
Group A-S	.129	.102	.061	.099

variances may have masked differences among the three changing probability groups. Individual differences probably were exaggerated by the presentation of individual rather than identical sequences of events to subjects within any one group.

Possibility of Ceiling Effect

As illustrated in Figure 5, average overconfidence for the three changing probability groups appears to increase during early trials, and appears to decline later for Groups S and A-S, but not for Group A. Figures 3 and 4 are suggestive of a ceiling effect for Groups S and A-S. Average OP for Group S reached a high in the study of .858, and average OP for Group A-S reached .895. Average SP for these groups exceeded .90 at some points, and reached a high of .983 in the fifth block for Group A-S. Overconfidence during later trials for these two groups may have been depressed by a ceiling effect. A decline in overconfidence at higher OP levels in a skill task reported by Howell (1972) is suggestive of a ceiling effect in that instance as well.

Factors Contributing to Overconfidence

As noted, overconfidence was found in the three changing probability groups, and underconfidence was found in the one unchanging group. Overconfidence probably resulted from several factors: information about the existing proportions in the underlying population as provided by draws, information about the changes in the underlying population as a result of the draws, and the general atmosphere of rising probabilities. The significant main effect of blocks found in the Groups X Blocks ANOVA for Phase I while no main effect of groups was found suggests that the atmosphere of rising probabilities present in all three changing probability groups was a factor contributing heavily to subjects' overconfidence. This notion is supported by the fact that subjects in all groups seemed to revise their SP estimates in a similar way, in spite of the fact that the different sampling schemes employed insured that identical samples drawn in each of the three groups had different consequences for the different groups. Increases in relative proportions of the focal event color resulted from draws of that color in Group A, from draws of the non-focal color in Group S, and from any draw in Group A-S. However, subjects in the different groups seemed to handle the information obtained through samples in similar fashion. Although it might be expected that subjects in Group A would increase estimations of SP after 5 draws in which to focal event color predominated (3, 4, or 5 focal event chips out of 5 draws) to a greater extent than would subjects in Group S, since a greater proportion of focal event chips led to a greater change in the underlying population for Group A, this expected difference was not found. In fact, there seemed to be little if any difference between groups in the average revision of SP after obtaining equal numbers of focal event chips in the 5 draws immediately preceeding an estimation. Figure 6 illustrates the average SP revisions for each of the groups on the estimation immediately following draws of low proportions (0, 1, or 2 focal event chips out of 5 draws) and average revisions following draws of high proportions (3, 4, or 5 focal event chips out of 5 draws). Average revisions for all groups increased with greater proportions of focal event chips in the 5 draws immediately preceeding an estimate.

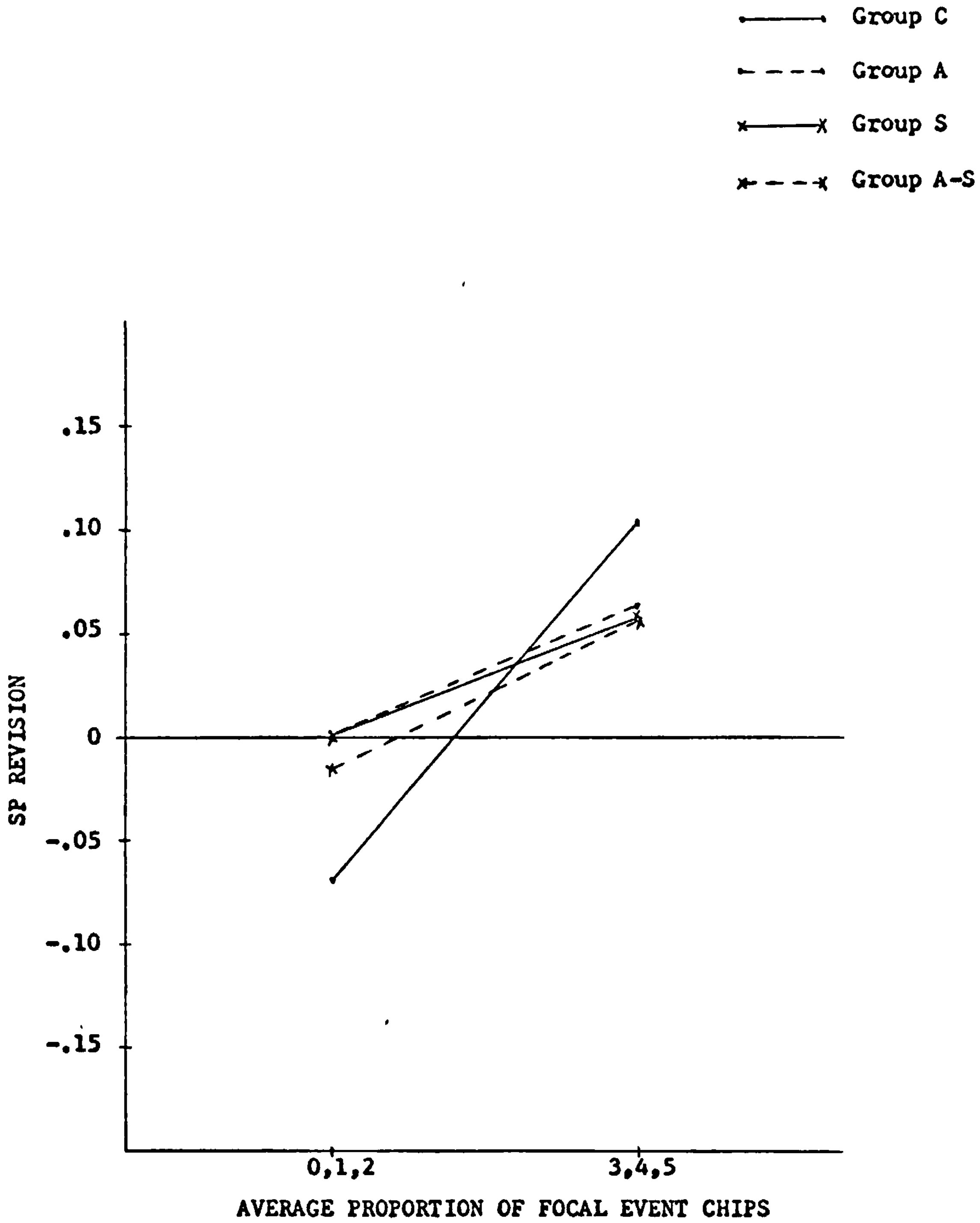

Figure 6. Average SP Revision following draws of 0, 1, 2 or draws of 3, 4, 5 focal event chips out of the five draws immediately preceeding the estimation; Phase I.

The curves for the three changing probability groups are nearly identical. It is noted, however, that this graph is not without confounds: draws of 0, 1, or 2 focal event chips occurred more frequently in early trials than in late trials for the changing probability groups, while the opposite was true for draws of 3, 4, or 5 focal event chips. Since average SP increased with trials, a ceiling effect might have depressed revisions of SP after 3, 4, or 5 focal event draws, and such a ceiling might have had greater effect for some groups than for others.

Underconfidence in Group C

The underconfidence in Group C must be considered. It could be argued that if underconfidence is associated with an event of low OP (such as drawing the focal event color in Group C), then overconfidence should be associated with an event of high OP (such as drawing the non-focal color in Group C). In a study of subjects' judgments of proportion in sequences of binary events, Pitz (1966) found constant errors of overestimation of proportions greater than .50, and underestimation of proportions less than .50. Since Pitz' task and proceedure differed from that of the present study, and since the sequence Pitz used were of both stable and changing proportions, it cannot be determined whether this overestimation would occur when subjects were estimating the probability of occurrence of an event of high probability when the sequences used reflected a constant underlying population. If such overestimation were to occur, however, it would not account for all overconfidence found in the present study, since overconfidence was found in all three changing probability groups at OP levels well below .50.

In Group A, average OP did not exceed .50 until the eleventh estimation trial, and reached a high on the final trial in Phase I of only .582. In connection with the underestimation of low proportions and overestimation of high proportions reported by Pitz, it is important to note that Howell (1972) explored a range of OP levels, and found overconfidence of success in skill tasks at high and low OP.

Effect of Shift from Phase I to Phase II

A Groups (A, S, and A-S) by Phases (I, II) ANOVA was performed to assess the effect of the shift from changing to stable probabilities. In this ANOVA the average of the last two trial blocks in Phase I was campared to the average of the two trial blocks in Phase II. The last two blocks (Blocks 4 and 5) of Phase I were selected in order to compare the behavior of subjects just before and just after the shift. A main effect of phases ($F_{1, 51} = 7.628$, $p < .01$) and an interaction of groups and phases ($F_{2, 51} = 3.492$, $p < .05$) were found. (See Table 5 for a summary of the Groups X Phases ANOVA.)

The main effect of phases found in the Groups X Phases ANOVA can be seen in graphic form in Figure 5. Overconfidence for all three changing probability groups was less in Phase II than in Phase I, although the Student's t test found overconfidence in both phases significantly different from zero. It appears that some lingering overconfidence was carried over into the stable probability phase (Phase II) from the changing probability phase (Phase I). There are difficulties with this interpretation, however, since there is some uncertainty surrounding the possibility of overestimation of proportions greater than .50 in a

Table 5. Groups (A, S, and A-S) by Phases (I and II) Analysis of Variance.

SOURCE OF VARIATION	SS	df	MS	F	
Between Subjects	1.365436	53			
Groups	.141324	2	.070662	2.944	ns.
_Ss_ Within Groups	1.224112	51	.024002		
Within Subjects	.669766	54			
Phases	.077870	1	.077870	7.628	**
Groups X Phases	.071300	2	.035650	3.492	*
Phases X _Ss_ Within Groups	.520596	51	.010208		

* $p < .05$

** $p < .01$

chance task of stable proportions, as reported by Pitz (1966).[2] It is impossible to determine whether the overconfidence in Phase II was greater than that which might have been expected in a chance task of equal OP which had not been preceeded by a changing probability phase. It does seem likely that the overconfidence in Phase II in the present study was greater than might have been expected in a one phase stable proportion situation for Group A, at least, since the average OP during Phase II for this group was only .582, while average overconfidence was .106. If there was any real carry over effect in Group A, it is doubtful that such an effect would have persisted over a large number of trials, since overconfidence for Group A in Phase II appears to decline.

Howell (1972) reported overconfidence in the asymptotic stage of the skill task. However, that asymptote and the asymptote in the present study are not strictly comparable, since there is no evidence that the subjects in the Howell study were aware that they had, in fact, reached asymptotic performance, while subjects in the present study were informed that all chips drawn would be replaced during Phase II, and this replacement was carried out in their presence. Therefore, subjects in the Howell study might still have expected to encounter further improvement, while those in the present study would not. It seems likely that a greater magnitude of overconfidence during the asymptotic stage could be attributed to a carry over effect in the Howell study than in the present study. Upon entering Phase II, subjects in the present study could have approached the estimation of current probabilities as a new task, using only those samples drawn during Phase II as a basis for estimations. Although one subject specifically stated that this was his approach, it is not known whether any other subject employed it.

The F ratio in the Groups X Phases ANOVA for a main effect of groups approaches significance at the .05 probability level. When the last two blocks of Phase I are combined with Phase II, average overconfidence for Group A is higher than that for Groups S and A-S.

Conclusions

The present study demonstrates that overconfidence can occur in a chance task as well as in a skill task, and indicates that the presence of changing probabilities may be a factor contributing to overconfidence in skill tasks. Since changing probabilities exist in the real world in both internally and externally controlled situations, further investigation of subjective probabilities in changing probability situations is indicated. Psychological studies of external and internal control have tended to employ stable probability tasks and increasing probability tasks (or tasks to which the notion of increasing probabilities might have been generalized), respectively. This fact might help to explain the apparant tendency for the internal tasks studied to result in overconfidence.

Suggestions for Future Research

The increasing probabilities of success in a skill task are generally viewed, at least in part, as a function of the state of the performer. Factors over which he has some control, i.e., practice, attentiveness, etc., contribute to success. Errors, or lack of success, in a skill task are similarly attributable to changeable states such as fatigue, inattention, or too little practice. Probabilities of success in skill tasks are not viewed as stable, since they are at least partially dependent upon changeable states of the subject. Since skill tasks are not viewed as stable, and since overconfidence has been found in both skill and chance tasks in which probabilities of success or of the focal event increased, future research in the area of changing probabilities should include studies in which those probabilities decrease. It is suggested that the use of a chance task in which probabilities of the focal event decrease would result in underestimation of those probabilities at both high and low objective probability levels. If subjects are consistently overconfident of success in skill situations, as suggested by Howell, then they should overestimate the probabilities of success in a situation in which those probabilities decrease, and should overestimate to the same degree as do subjects in skill tasks in which probabilities of success increase. If, however, the overconfidence found in skill tasks is partially due to the presence of increasing probabilities of success, then in a situation in which probabilities of success decrease, subjects should overestimate those probabilities to a lesser degree than should subjects in the increasing probability situation, and might even underestimate in the decreasing probability situation. Since a skill task in which probabilities of success decrease can be viewed as one in which probabilities of failure increase, some subjects might be asked to estimate probabilities of failure rather than probabilities of success. A skill task which becomes increasingly difficult with fatigue might be employed to provide a skill task in which probabilities of success decrease while probabilities of failure increase.

In the changing probability tasks under consideration in the present study, subjects were aware of the increasing nature of the probabilities. Increasing or decreasing probabilities in a task in which subjects were not informed of the changeable probabilities might result in over- or underestimation, respectively. The present study indicates that an atmosphere of change may contribute to systematic over- or underestimation. If subjects who were aware of the increasing or decreasing nature of the task consistently over- or underestimated to a greater degree than did subjects who received identical sequences but who were not informed of the changeable probabilities, it would lend support to the notion that the atmosphere of changing probabilities is an important factor in over- or underconfidence. This could be carried further by informing some subjects that change would occur, but by presenting these subjects with sequences which reflected stable underlying proportions. The appropriate over- or underestimation might also be found in this instance, although the information actually obtained from the sequences would not suggest change in probabilities.

Shifting from changing probabilities to unchanging probabilities without informing subjects of such a shift should result in continued over- or underestimation, which might or might not persist over a number of trials. This situation would be more closely comparable with the asymptotic stage of the Howell studies than was the situation in the present study. This shift could be accomplished by using two urns, as discussed by Restle and Greeno (1970, pg. 3) in connection with their learning models. One urn, the subject's urn, would contain a low proportion of the focal event, while the other urn, the experimenter's urn, would contain a relatively high proportion. The subject would draw from his urn, and replacements for these draws would be dictated by draws from the experimenter's urn. Replacements for the subject's urn would be taken from a supply for that purpose, and the draws from the experimenter's urn would be replaced in that urn so that the population in the experimenter's urn would remain constant. This method has the advantages of keeping the size of the population in the subject's urn constant, of allowing for slower rates of change in the subject's urn, and of providing an asymptote which could be set at any OP level desired without requiring a change in the sampling scheme. This task could be varied to result in decreasing probabilities. A disadvantage connected with this methoed lies in the fact that the asymptote would be equal to the proportion in the experimenter's urn, and subjects might well recognize this fact and refuse to make estimations beyond their estimates of this proportion. This would result in a ceiling effect. In order to avoid a ceiling of any level less than the necessary ceiling imposed by the probability range of 0—1, it may be necessary to resort to an apparatus which would allow for changes to be made without the subjects' knowledge so that deception might be carried out.

Changing probability tasks could be varied in context. For example, it is possible that subjects estimating probabilities of success in a skill task in which they were personally involved might overestimate to a greater degree than would subjects estimating probabilities of success for another individual. Subjects estimating probabilities of success for another individual might also vary their estimations in relations to the characteristics attributed to that other individual.

Pitz (1966) found a tendency for overestimation of events whose proportions were greater than .50 to be slightly greater when the number of events between judgments was small. That is, when judgments were revised after each event, overestimation was found to be greater than when judgments were taken only at the end of the entire sequence or at intervals of 5 or 10 events. It would be wise, therefore, to include size of the interval between judgments as an independent variable in some study in order to assess the effect of differences in the number of events making up the interval.

BIBLIOGRAPHY

BIBLIOGRAPHY

Feather, N. T. Success Probability and Choice Behavior. Journal of Experimental Psychology, 1959, 58(4), 257-266.

Howell, W. C. An Evaluation of Subjective Probability in a Visual Discrimination Task. Journal of Experimental Psychology, 1967, 75(4), 479-486.

Howell, W. C. Uncertainty from Internal and External Sources: A Clear Case of Overconfidence. Journal of Experimental Psychology, 1971, 89(2), 240-243.

Howell, W. C. Compounding Uncertainty from Internal Sources. Journal of Experimental Psychology, 1972, 95(1), 6-13.

Peterson, C. R., and Beach, L. R. Man as an Intuitive Statistician. Psychological Bulletin, 1967, 68(1), 29-46.

Peterson, C. R., and DuCharme, W. M. A Primacy Effect in Subjective Probability Revisions. Journal of Experimental Psychology, 1967, 75(1), 61-65.

Pitz, G. F. The Sequential Judgment of Proportion. Psychonomic Science, 1966, 4(12), 397-398.

Restle, F., and Greeno, J. G. Introduction to Mathematical Psychology, Reading, Massachusetts: Addison-Wesley Publishing Company, 1971, 1-23.

Winer, B. J. Statistical Principles in Experimental Design. New York, New York: McGraw-Hill Book Company, 1962, 298-318.

FOOTNOTES

FOOTNOTES

1. Researchers working in the area of Bayesian statistical inference have studied subjective probabilities in chance tasks with changing probabilities. However, the typical Bayesian task requires subjects to supply subjective probabilities concerning the likelihood that the sequences presented are drawn from one of two possible populations (e.g., Peterson and DuCharme, 1967). The proportions in the samples are varied, and the SP studied is the probability that the population sampled is Population A or Population B. This is quite different from a situation in which subjects are required, not to choose between two alternative hypotheses for which proportions are given, but to formulate their own hypotheses about the proportions of the underlying population.

2. Peterson and Beach (1967) review the experimental research using probability theory and statistics as a framework within which to study human statistical inference. These authors observe that when subjects have estimated proportions of binary events "the relation between mean estimates and sample proportions is described well by an identity function" and that "the deviations from this function are small; the maximum deviation of the mean estimate from the sample proportion is usually only .03—.05, and the average deviation is close to zero".

APPENDIX

Illustrations of SP, OP, and SP-OP by Estimation Trials

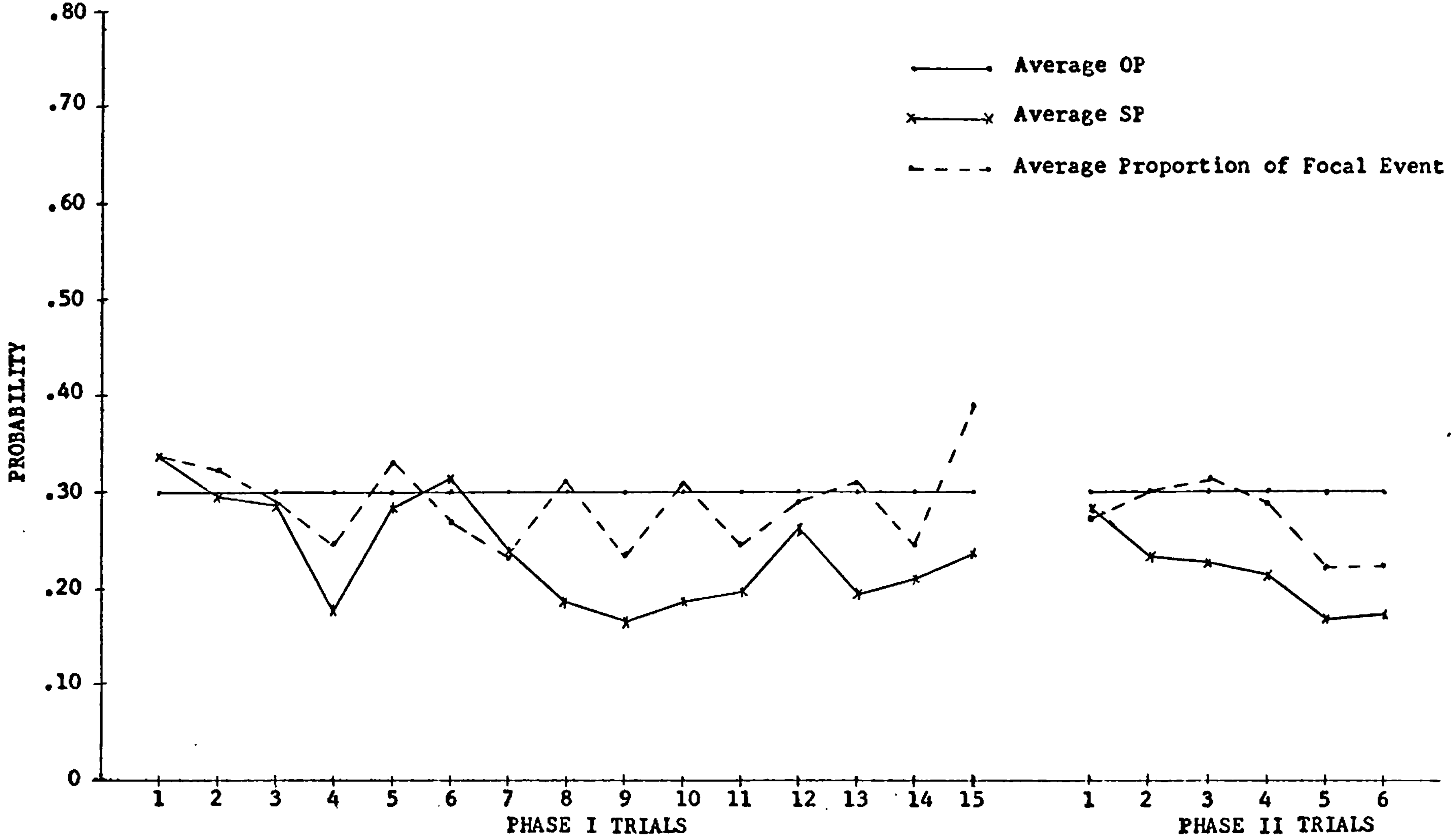

Figure 7. Average OP, Average SP, and Average Proportion of Focal Event Chips drawn in the five draws immediately preceeding an estimation; Group C.

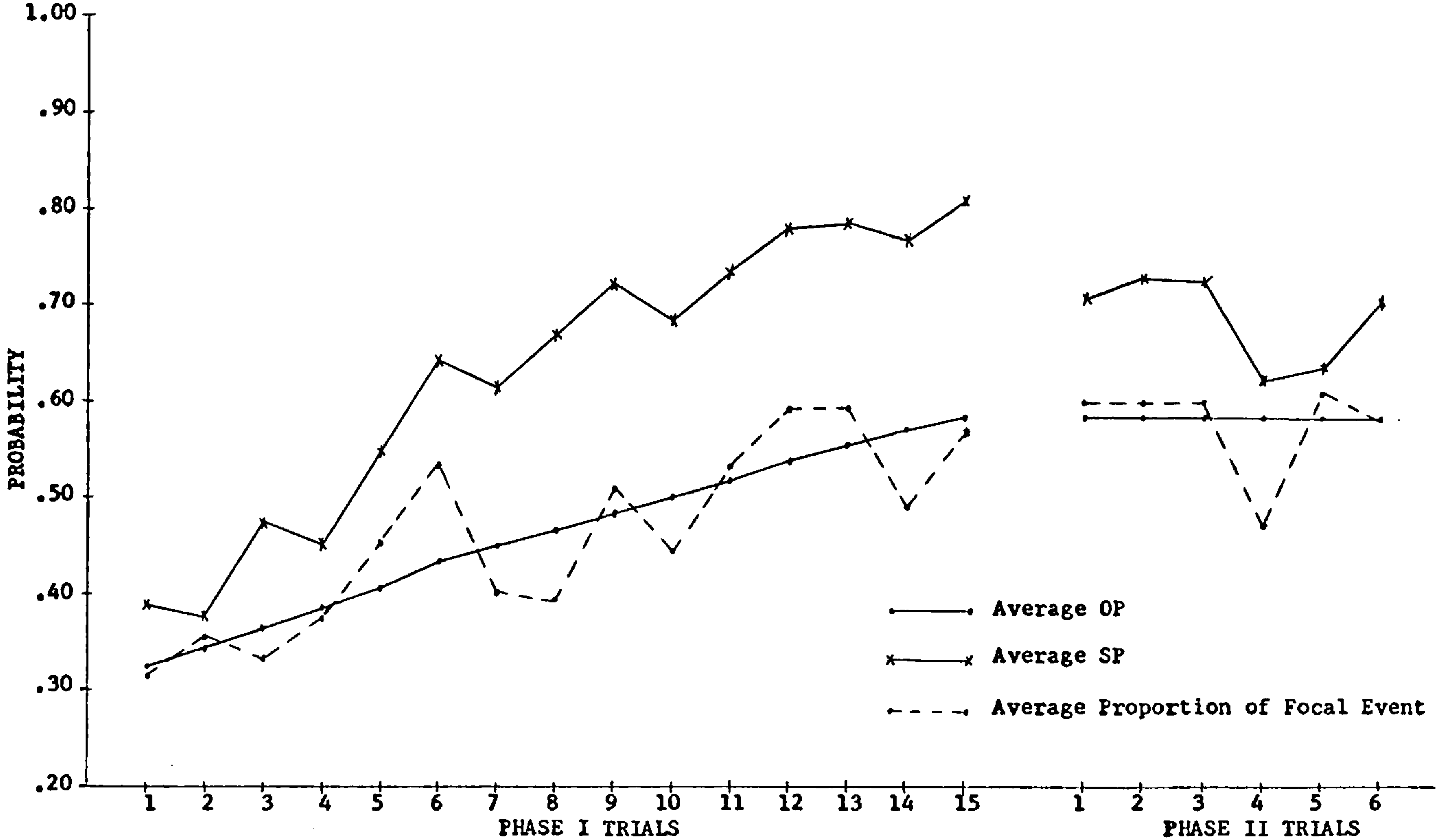

Figure 8. Average OP, Average SP, and Average Proportion of Focal Event Chips drawn in the five draws immediately preceeding an estimation; Group A.

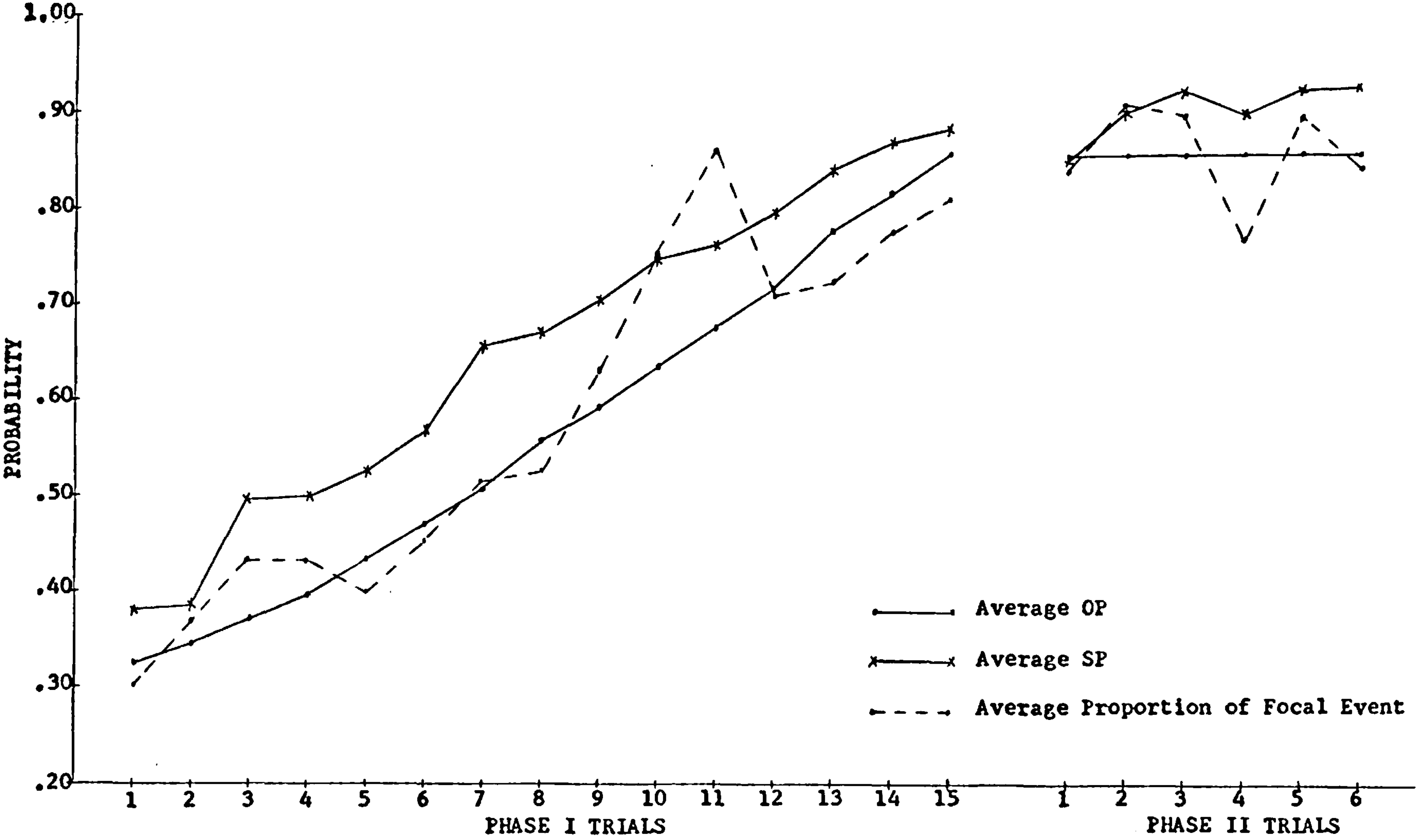

Figure 9. Average OP, Average SP, and Average Proportion of Focal Event Chips drawn in the five draws immediately preceeding an estimation; Group S.

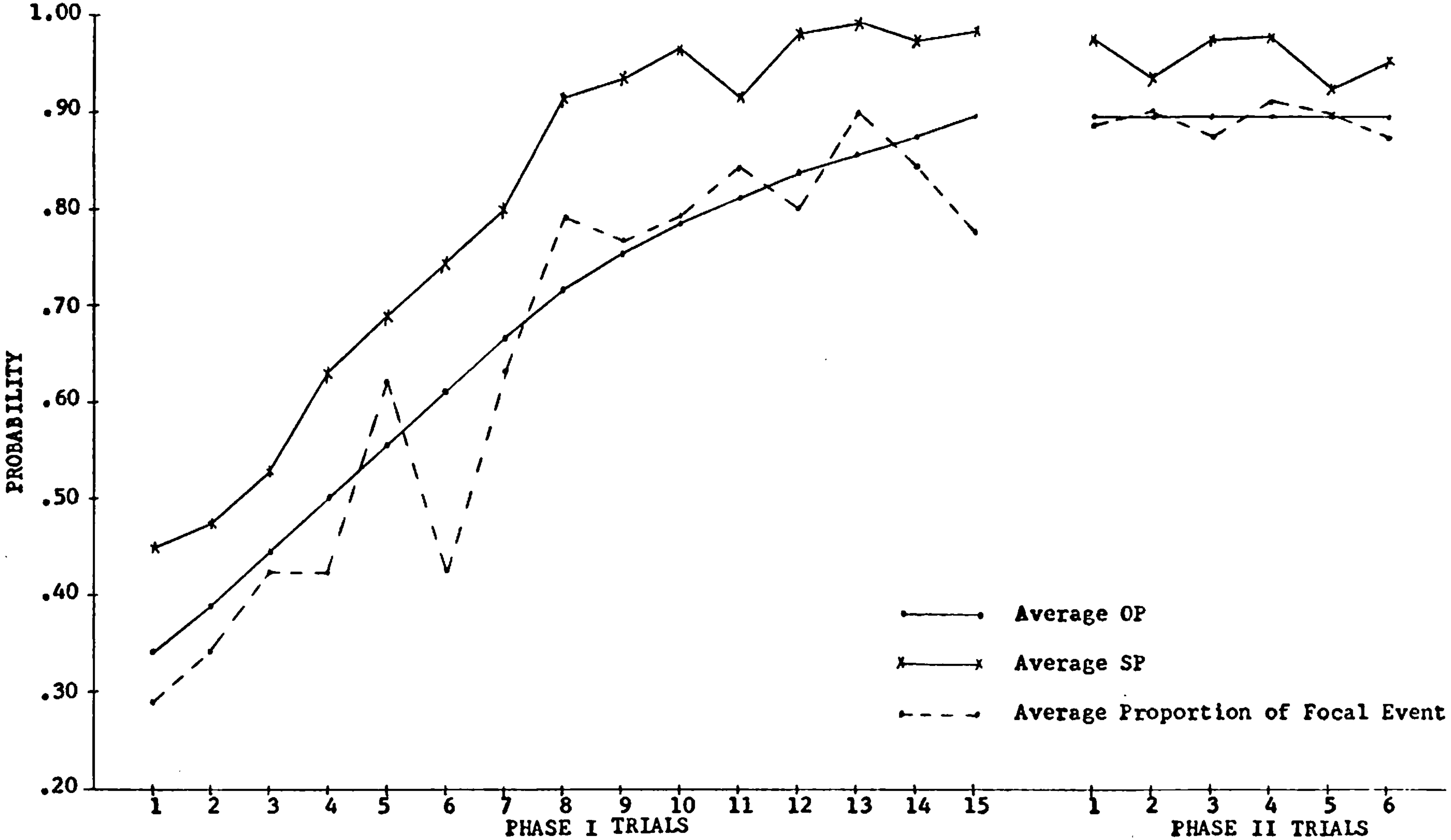

Figure 10. Average OP, Average SP, and Average Proportion of Focal Event Chips drawn in the five draws immediately preceeding an estimation; Group A-S.

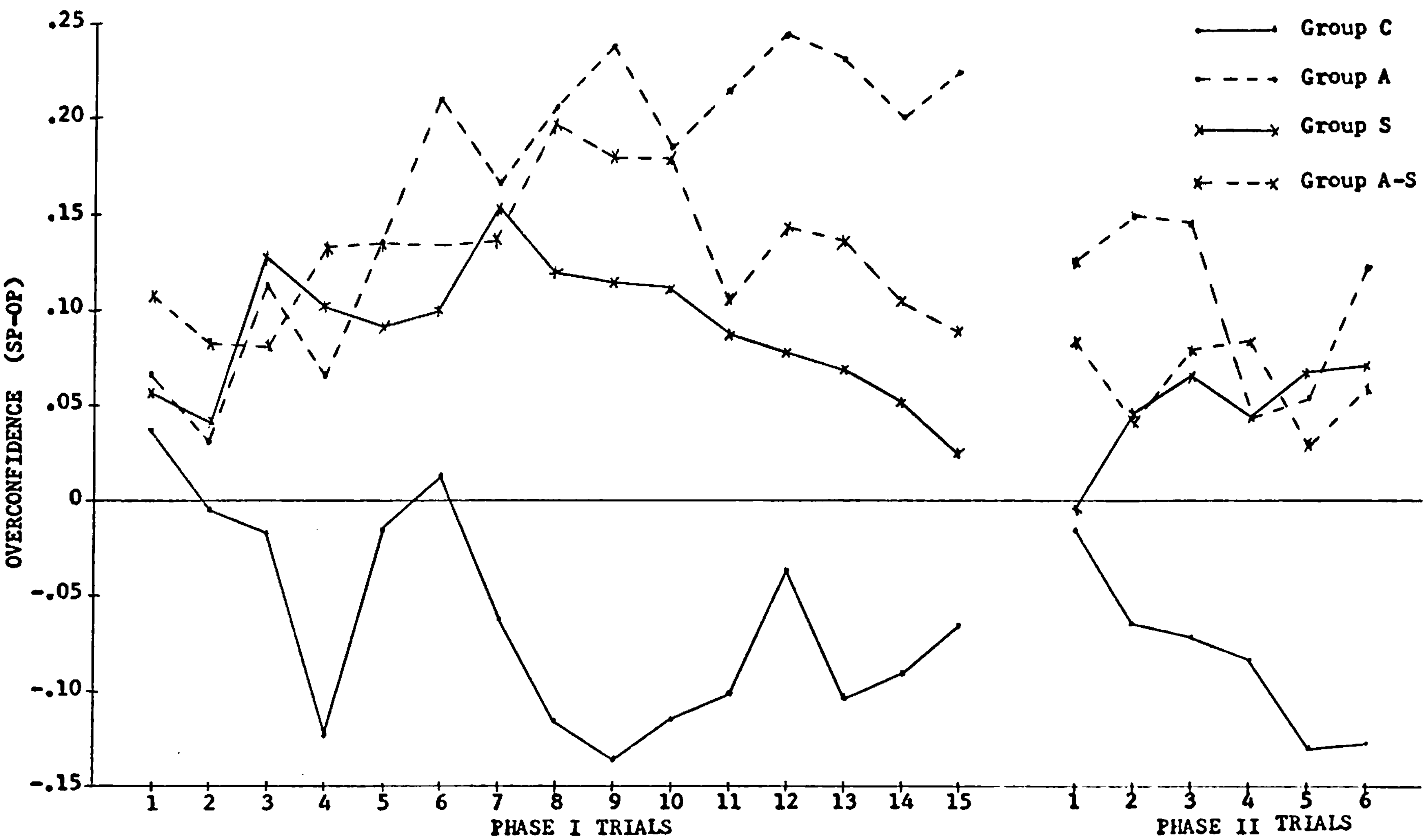

Figure 11. Average Overconfidence (SP-OP).