Br. J. Psychol. (1975), **66**, 3, *pp.* 357–372
Printed in Great Britain

EXPERIMENTER BIAS AND SUBLIMINAL PERCEPTION

By PAUL J. BARBER
Birkbeck College, University of London

AND J. PHILIPPE RUSHTON*
London School of Economics

It has been suggested that subliminal perception phenomena may be in part due to experimenter bias effects. Two studies that obtained positive evidence of subliminal perception were therefore replicated with experimenters tested under blind and not blind conditions. There was only marginal support for the subliminal perception hypothesis and, although there were fairly clear indications of diffuse experimenter effects, the evidence for the experimenter bias explanation of subliminal perception was not strong. The need for more extensive replications of subliminal perception researches is noted. It is argued that the experimenter bias hypothesis lacks detail and generality; it is essential for it to be examined in the context of theoretically substantial issues.

It was suggested by Neisser (1967) that subliminal perception effects may be mediated by 'demand characteristics'. This would be a serious assertion if it were more specific, and if it had an empirical basis. These studies examined a strong version of Neisser's hypothesis: that subliminal perception effects are mediated by experimenter bias (Rosenthal, 1966, 1969). A direct approach was adopted, replicating 'successful' paradigms with and without an experimenter-blind condition.

Two studies were selected which Neisser (1967) had mentioned in some detail, and which otherwise seemed quite well-known and were associated with strong and explicit theoretical predictions. The studies chosen for replication were those of Smith *et al.* (1959) and Spence & Holland (1962). Both were implemented with and without an experimenter-blind condition, and so it was possible to test both the subliminal perception hypothesis and the experimenter-bias hypothesis.

The most extreme form of the present experimenter-bias hypothesis states that there will be no subliminal perception effect except when, in some relevant sense, the experimenter is not blind (i.e. he can bias). According to the subliminal perception hypothesis the subliminal perception effects will occur both when the experimenter is blind and when he is not blind. There is of course room for a compromise hypothesis maintaining that the subliminal perception effects occurring when the experimenter is blind will be enhanced when he is not blind.

Experiment I

Spence & Holland (1962), arguing within a psychoanalytic framework, hypothesized that reduced awareness would lead to a diffusion of associative activity. In particular, they reasoned that a subliminal stimulus would be registered and prime an associative network in which it would produce more widespread and diverse activity than if it were presented supraliminally. Hence their experiment: the subject was presented with the word CHEESE either subliminally, supraliminally or not at all.

* Present address: Department of Psychology, York University, Toronto, Canada.

He was then given a free-recall trial of a word list comprising associates and non-associates of the word CHEESE. Recall of CHEESE associates was facilitated only in the subliminal condition.

The present study involved a replication of this experiment using an individual test procedure in preference to the group testing of the original. Each assistant experimenter (E_A) administered the experiment under blind and not blind conditions.

Method

The experiment extended over two days for each E_A, who tested six subjects under each of the three cue conditions used by Spence & Holland (1962): subliminal, blank and supraliminal. Half of the subjects in each group were tested with E_A blind with respect to the cue condition. There was no simple way of preventing E_A from knowing when the subject was presented with a supraliminal cue, so the experimenter-blind manipulation strictly refers to the subliminal–blank distinction.

Individual word recognition thresholds were estimated prior to the cue treatment. Following the cue treatment, the Spence & Holland CHEESE word list was presented and the subject gave immediate written recall of this material.

Experimenters and subjects

The E_As were two male and two female undergraduate psychology students, familiar with subliminal perception evidence and expressing a belief in the validity of the phenomena. Each E_A found and tested 18 subjects aged between 18 and 60. In all there were 72 subjects (42 male, 30 female; mean age 27 years).

Procedure

Experimenter procedure. The initial briefing of the E_As emphasized that subliminal perception effects had been demonstrated effectively enough, but that it had been alleged that experimenter effects might overlay the findings. Therefore it was the objective of the present study to replicate the original finding before conducting further experiments based on this paradigm. The E_A was then advised that the CHEESE recall effect had been demonstrated clearly and beyond doubt and the effect was described in some detail to be sure that it was quite familiar. It was expected, wrongly as it turned out, that the E_As would know about Neisser's (1967) demand characteristics argument about subliminal perception. It was in any event feasible that E_As would infer the experimenters' concern about experimenter effects from the design of this study, and it was for these reasons that there was no attempt in Expt. I to conceal this concern. Nevertheless, the briefing was slanted to give E_A the impression that the primary objective was to replicate the subliminal perception effect and that experimenter effects were merely an inevitable nuisance factor. This should not have eliminated positive bias effects except of a deliberate kind.

The experimenter, entirely concealed from E_A and the subject by a large screen, was responsible for the operation of the tachistoscope. The subjects did not know that the experimenter was behind the screen. The recall protocols were sealed in envelopes by E_A as soon after recall as possible. No data were inspected until the experiment was completed and both experimenters scored all the data independently.

Preliminary threshold estimation. The recognition threshold for each subject was obtained for each of six words equated with CHEESE for length and word frequency. The words were typed in capitals on grey cards and were presented in a Cambridge Model tachistoscope with a plain white card in the inspection field. The simple ascending method of limits was used. The critical exposure duration was taken as four duration levels (generally 40 msec.) below the lowest exposure time obtained during this preliminary session.

Cue treatment. In the subliminal and blank conditions the respective cues were the word CHEESE or the symbols XXXXXX. The cue was presented tachistoscopically five times at 5 sec. intervals at the critical exposure duration determined in the preliminary session. Multiple presentations of the cue were used to reduce the chance that the subject inadvertently failed to receive the cue (e.g. because he blinked coincidentally with the presentation). Although repeated tachistoscopic

Table 1. *Analysis of variance of recall data: Expt. I*

Source	D.F.	M.S.	*F*
Between subjects	71		
Bias (A)	1	6·25	1·49
Cue (B)	2	0·88	0·27
Experimenters (E)	3	2·43	0·51
A × B	2	2·15	1·13
A × E	3	4·19	0·88
B × E	6	3·28	0·69
A × B × E	6	1·90	0·40
Subjects within groups	48	4·74	
Within subjects	72		
Recall category (C)	1	53·78	28·75*
A × C	1	0·25	0·08
B × C	2	5·34	3·66
E × C	3	1·87	0·69
A × B × C	2	1·27	0·33
A × E × C	3	3·16	1·16
B × E × C	6	1·46	0·54
A × B × E × C	6	3·82	1·40
C × subjects within groups	48	2·73	

* $P < 0{\cdot}05$.

presentations increase the probability of word recognition (Haber & Hershenson, 1965) it is unlikely that the effect is important for a duration as far below threshold as those used here (Haslerud, 1964). In the supraliminal condition CHEESE was presented four times at 5 sec. intervals at an exposure duration of 1 sec. In all cases the subject was instructed to say if possible what was being presented.

Recall procedure. Following immediately after the cue treatment, the subject was presented with a deck of 20 plain playing cards on which were typed in capitals the words from the Spence & Holland list (excepting their primacy and recency buffer items since a fixed order was not used in this replication). These words were: (*a*) CHEESE associated: COW, BREAD, CAVE, MOON, COTTAGE, GREEN, BRICK, SMELL, MOUSE, SOUR; (*b*) control words: FLAG, ARCH, CHAIR, REST, GIFT, FRAIL, SAND, HEDGE, MINOR, TRUNK. The order of the recall list was produced by E_A by shuffling the deck thoroughly for each subject. The presentation of the list was controlled by the subject but was, following a demonstration by E_A, intended to be at about one word per second.

Results

In the source study Spence & Holland used differential recall of CHEESE associates as the dependent variable. This was defined as the difference between the number of CHEESE associates and the number of control words recalled. Several versions of the recall scores were examined in the present study and the conclusions from them were similar in all essential respects. The analysis of variance summarized in Table 1 is typical of the pattern of statistical significance (the same in all analyses) and of the magnitude of the effects. This pattern also emerged if the supraliminal condition (for which the blind treatment was suspect) was excluded.

The only significant effect on recall was that corresponding to the difference between CHEESE associates and control words ($F = 28{\cdot}75$; d.f. = 1, 3; $P < 0{\cdot}05$). This can be seen in Table 2, which presents mean recall scores averaged across E_As.

An overall parametric analysis of variance was preferred to the piecemeal non-parametric treatment that otherwise would be necessary to examine the data in all

Table 2. *Mean recall averaged across assistant experimenters: Expt. I*

Bias condition	Cue	Recall category: Cheese associates	Recall category: Control words
Blind	Subliminal	5·67	3·50
	Blank	4·67	3·67
	Supraliminal	4·17	3·92
Not blind	Subliminal	5·42	3·83
	Blank	5·25	3·75
	Supraliminal	5·33	4·50

the relevant aspects. It was of course essential that this analysis should allow the same basic question to be put as in the source study.

The main finding of the source study was not replicated, since the cue × recall category (representing Spence & Holland's differential recall variable) was not significant at the 5 per cent level ($F = 3{\cdot}56$; d.f. = 2, 6; $0{\cdot}10 > P > 0{\cdot}05$), although it just attained significance at the 10 per cent level. Entirely crucial to the experimenter bias hypothesis is the cue × recall category × bias interaction. This also was not significant ($F = 0{\cdot}33$; d.f. = 2, 6; $P > 0{\cdot}05$), nor was any other interaction which might qualify the effect of these variables. There were no significant interactions with the experimenter factor.

The analysis of variance summarized in Table 1 was constructed on the assumption that experimenters should be treated as a random factor (using interactions with experimenters as error terms for fixed main and interaction effects, but using interactions with subjects otherwise). The intention was to aim for generality in this way, but it makes no difference to the above conclusions if experimenters are (improperly) treated as a fixed factor. Nor is anything gained by the dubious device of pooling the interactions involving experimenters to form a common error term; for instance, the cue × recall category interaction becomes statistically even less remarkable ($F = 1{\cdot}98$; d.f. = 2, 55; $P > 0{\cdot}10$). A planned comparison (Kirk, 1968) of the means of the subliminal condition against the combined means of the supraliminal and blank conditions approached significance ($F = 5{\cdot}31$; d.f. = 1, 6; $0{\cdot}1 > P > 0{\cdot}05$), suggesting a marginal tendency for the recall of CHEESE associates to be additionally facilitated when the cue word was presented subliminally.

An experimenter effect was found in the preliminary threshold data. An analysis of variance of the logarithms of the critical exposure durations recorded for each subject indicated that the experimenter factor was significant ($F = 4{\cdot}864$; d.f. = 3, 48; $P < 0{\cdot}01$) and the experimenters × bias × cue interaction was significant ($F = 2{\cdot}371$; d.f. = 6, 48; $P < 0{\cdot}05$). In this last case there was no discernible pattern, although there was little variability in thresholds when E_A was blind. This is reassuring since it suggests that E_As' initial treatment of their subjects was, as intended, independent of the cue condition; this is consistent with the notion that E_As did not receive any useful information from the concealed experimenter when the blind condition applied. The overall mean preliminary threshold was 135 msec.

Experiment II

In an attempt to exaggerate any effect there might be due to the expectations of the experimenter, one of the E_As from Expt. I tested further subjects, ostensibly under the subliminal and blank cue conditions. In fact, half of the subjects were

Table 3. *Mean recall scores and the expectations of the assistant experimenter: Expt. II*

		Recall category	
Expected cue	Actual cue	Cheese associates	Control words
Subliminal	Subliminal	5·10	3·90
	Blank	4·30	3·70
Blank	Subliminal	4·20	3·50
	Blank	5·30	3·20

tested, unknown to the E_A, with the opposite cue treatment to that which she expected. In this way, it was possible to examine the effects of the cue and of the E_A's expectancy in combination and in opposition.

Method

Experimenter and subjects

One of the female E_As from Expt. I was employed to contact and run 40 further subjects in a subset of the conditions of the first study. The E_A was not asked to record details of subjects but to ensure that they were in the same age range as before (18–60 years).

Procedure

Experimenter procedure. The E_A was told that in Expt. I her results in the subliminal *v.* blank condition were the only ones to attain statistical significance and that it was now intended to investigate the reliability of the effects. She was asked to test 20 subjects under the subliminal CHEESE cue condition and another 20 subjects under the blank cue condition. Correct information was given about the cue for the first 10 subjects on each of the two extra experimental sessions required, but the cue was opposite to what E_A had been told for the second set of 10 subjects in each session. The E_A was debriefed at the end of data collection.

Preliminary threshold estimation and post-recall check. The preliminaries for each subject were the same as in Expt. I. After the recall test the subject was asked to indicate on a seven-point scale the degree of certainty with which he judged the pre-recall subliminal cue to be CHEESE. This was intended to provide an indirect check on whether, in the event that E_A was able to bias, the subject was able to detect this bias. It incidentally acted as a secondary check on the subliminality of the cue. Most subjects (34/40) showed uncertainty about the cue (rating 4) and the distribution of scores was so even that the data were not analysed further.

Cue treatment and recall procedure. The pre-recall treatment for a given subject consisted of either the subliminal or blank condition of Expt. I. The recall procedure was unchanged from Expt. I.

Results

Conclusions for the various recall indices (absolute or relative scores) were the same and absolute recall scores are reported here (means are presented in Table 3).

As in Expt. I only recall category contributed significantly to the total variance; again recall of CHEESE associates was reliably better than of control words. If the E_As were not biasing and there was a subliminal perception effect alone, then the cue ×

recall category interaction would be significant; this was not so ($F = 0{\cdot}35$; d.f. = 1, 36; $P > 0{\cdot}05$). On the other hand, if there was no subliminal perception effect but E_A could bias, then the expectancy × recall category interaction would be significant; this did not occur ($F = 0{\cdot}54$; d.f. = 1, 36; $P > 0{\cdot}05$). A joint hypothesis that there would be both subliminal perception and experimenter bias effects was not supported; the expectancy × cue × recall category interaction was not significant 2·16; d.f. = ($F = 1$, 36; $P > 0{\cdot}05$).

Table 4. *Comparison of source and replication studies: Expt. I*

Basis of comparison	Source study (Spence & Holland, 1962)	Replication study
Subjects	$n = 59$	$n = 72$
Source of subjects	Undergraduate classes	Selected by E_A locally in university precinct
Procedure	Group-testing; and subliminal and blank groups run together	Individual
Display	Slide projector + shutter	Tachistoscope
Exposure duration	$\frac{1}{100}$ sec. and a post-experimental check	Individually determined. Range 30–100 msec.
Blank cue condition	Subjects looked away from screen	Subliminal presentation of XXXXXX
Recall list	Word list of 10 CHEESE associates + 10 control words + three primacy and three recency 'buffer' items	Word list of 10 CHEESE associates + 10 control words (same as source)
Presentation order of recall list	Fixed	Random
Presentation modality of recall list	Auditory (by experimenter)	Visual (by subject)
Experimenters	'Naive instructor'	Four assistant experimenters + experimenter behind screen to operate display

Preliminary discussion of Experiments I and II

Procedural differences between the source study and the experiments reported above may account for the failure to replicate. On balance, the advantage in terms of sensitivity to subliminal perception effects lies with the replication study. Thus the use of an individual-testing procedure was an improvement over the original group-testing approach as exposure durations could be used that were more likely to be appropriate for a given individual, and it was easier to monitor the subject's attention to the task. Other differences (see Table 4) seem to be the kinds of variation to which the phenomena would be insensitive. The procedural differences therefore favoured the replication study as a detector of subliminal perception effects.

The failure to replicate cannot readily be attributed to experimenter bias in the original study since, although diffuse effects werefound on the preliminary thresholds, the blind *v.* not blind manipulation had no effect on the main dependent variables in the present study. Concerning the recall task the overall level of performance was on the whole better than in the original study, but this was probably due to the difference in effective list length; in the source study the primacy–recency

buffer items which were omitted in the replication were not included in the recall score.

An alternative explanation for the failure to replicate is that there was a negative experimenter effect (Dixon, 1971) in the present study; in some way the subjects were made unresponsive to the subliminal stimuli. This is not a convincing explanation because *inter alia* it runs counter to the expectations of both experimenters and E_As, but it is difficult to rule it out in this or any other study failing to find a subliminal effect.

In terms of statistical inference, either Spence & Holland committed a Type I error or a Type II error has been committed here. Again the balance seems to favour the replication study which used more powerful statistical tests, but comparison of power is not straightforward because the degrees of freedom on which the hypotheses were tested here were small (because the design was hierarchical). For this reason individual analysis, for each E_A considered separately, might be preferred, since the degrees of freedom would be more comparable to the original analysis. When this was done there was still no significant effect (other than recall category) for any E_A despite the increased degrees of freedom. Furthermore, as has been seen above, the use of pooled error terms did not lead to a different conclusion.

The other way in which the source and replication studies differ is in relation to the generality of the effect of recall category. In both replication experiments there was a clear superiority in recall of the CHEESE associates over the control words. In Spence & Holland's experiment this advantage was not statistically reliable in either of their blank and supraliminal conditions. The present findings of a general advantage to the associated words are in line with evidence on the relation between free recall and intralist associations (Deese, 1959; Rothkopf & Coke, 1961). In the light of this evidence it could be argued that Spence & Holland's findings were mediated not by the facilitative effect of the subliminal cue but by a depressive influence acting in the other conditions.

EXPERIMENT III

The second source study was the experiment by Smith *et al.* (1959). This experiment investigated the effects of the subliminal presentation of a word (HAPPY or ANGRY) alternating with a clearly visible line drawing of an expressionless face. The principal finding was that verbal descriptions of the face were judged subsequently to be significantly different on a dimension related to the subliminal cue. Apart from the inclusion of an experimenter blind *v.* not blind comparison, the major difference between the source experiment and the replication concerned the method of response recording, since in the replication the subject used a rating scale.

Method

Each subject rated a series of pictures of expressionless faces on a 'happy–angry' scale. Each picture was periodically interrupted by the subliminal presentation of one of the words HAPPY or ANGRY. The E_A either knew (not blind) or did not know (blind) which word was being used as the subliminal cue.

Stimulus material

Eight line drawings of faces rated as neutral by independent judges on a five-point HAPPY–ANGRY scale were used in the replication.

Experimenters and subjects

The E_As were one male and six female undergraduate psychology students who had expressed a conviction about the validity of the evidence on subliminal perception. Each E_A tested 24 subjects aged between 18 and 60. In consequence, 168 subjects (108 males, 57 females; unrecorded data for three subjects) participated in the experiment.

Procedure

Experimenter procedure. The E_As were advised of the general findings of the source experiment and it was stated that they would be replicating other workers' findings if they obtained subliminal perception effects. They were unaware of the experimenters' interest in experimenter effects. A more elaborate briefing was judged to be unnecessary because the E_As in Expt. I were found to be naive with respect to the demand characteristics hypothesis of subliminal perception. Each E_A attended for two sessions and the first was run with E_A not blind with respect to the subliminal cue word. At the end of this session E_A was instructed that in the second session he would not know what was the subliminal cue, so that a comparison of his performance, blind and not blind, could be made.

On the first session each E_A tested six subjects with HAPPY as the subliminal cue and six subjects with ANGRY as the cue. These were run in two blocks with the same cue within a block, and the order of blocks was alternated between E_As. Thus it was ensured that E_A would have a settled notion of what cue was being used. In the second session the cue word, which was taped into position by the experimenter, was randomly changed between subjects by the experimenter in the interval when E_A was fetching the next subject. Six subjects were tested with HAPPY as the cue, and six with ANGRY. The experimenter was in a nearby room and did not meet or communicate with E_A during the running of the experiment.

Pre- and post-experimental threshold procedure. Four words matched for length and word frequency with HAPPY and ANGRY were used in conjunction with an ancillary Cambridge Model tachistoscope to obtain the critical duration for the cue word. The ascending method of limits was used and the subliminal exposure duration was then taken as that level which was four duration steps below the lowest previously obtained threshold (averaging 87 msec.).

The threshold for each subject was again established at the end of the main task. If the post-experimental threshold was lower than or equal to the pre-experimental threshold, the subject was replaced; this was necessary for six subjects, evenly distributed among E_As. The words were typed in capitals on grey cards and the inspection field was a blank white card.

Cue treatment and mood rating task. The subliminal cue word was inserted by E_A in the not blind session, but was taped in position by the experimenter in the blind session. The cards bearing the line-drawings of faces were inserted (one at a time) in the inspection field of the tachistoscope, and on a given trial the subject was given 10–15 sec. to inspect the drawing. During this period E_A triggered a repeat circuit which presented the cue 10 times at the chosen critical duration at intervals of 0·5 sec.

The subject was asked to judge whether each face looked happy or angry, and he indicated his judgement on a card marked with a seven-point rating scale. There was a separate card for each face. The scale ranged from 1 for 'extremely happy', through 4 for 'neither happy nor angry', to 7 for 'extremely angry'. The order of presentation of the eight faces was separately randomized for each subject.

Results

The eight ratings for each subject were combined into two sets of four, corresponding to the first and second phases of the experiment, which were in fact undifferentiated as far as the subject was concerned. These pooled ratings were subjected to an

Table 5. *Analysis of variance of 'happy-angry' ratings: Expt. III*

Source	D.F.	M.S.	*F*
Between subjects	167		
Bias (A)	1	0·1163	0·04
Cue (B)	1	0·3127	2·37
Experimenters (E)	6	0·2767	1·04
A × B	1	0·0984	0·21
A × E	6	0·3181	1·21
B × E	6	0·1317	0·50
A × B × E	6	0·4626	1·75
Subjects within groups	140	0·2649	
Within subjects	168		
Phase (C)	1	0·0017	0·00
A × C	1	0·0672	0·11
B × C	1	0·1163	0·27
E × C	6	0·4787	1·42
A × B × C	1	0·6475	1·83
A × E × C	6	0·6310	1·87
B × E × C	6	0·4240	1·26
A × B × E × C	6	0·3528	1·05
C × subjects within groups	140	0·3366	

Table 6. *Mean ratings, averaged over assistant experimenters, on the 'happy–angry' scale: Expt. III*

		Cue condition	
Bias condition	Phase	Happy	Angry
Blind	First	4·27	4·25
	Second	4·26	4·33
Not blind	First	4·14	4·36
	Second	4·24	4·21

Table 7. *Mean ratings for male and female subjects classified by cue and bias conditions*

(Numbers of subjects per cell are shown in parentheses. MS_{error} = 0·126 with 157 d.f.)

		Cue	
Sex of subject	Bias condition	Happy	Angry
Female	Blind	4·11 (14)	4·31 (10)
	Not blind	4·35 (18)	4·19 (15)
Male	Blind	4·33 (27)	4·31 (30)
	Not blind	4·07 (24)	4·34 (27)

analysis of variance, summarized in Table 5. There were no significant main or interaction effects. This applied regardless of whether the experimenter factor was treated as random or fixed, and whether error terms were pooled or not. In particular, there was no subliminal perception effect (cue), nor an experimenter-bias effect (bias), nor an interaction between them. Furthermore, neither of these effects showed any sign of developing within a session (no interaction with phase).

The general impression of numerical calm can be seen in Table 6, which shows the mean ratings averaged over E_As.

Table 8. *Comparison of source and replication studies: Expt. III*

Basis of comparison	Source study (Smith *et al.*, 1959)	Replication study
Subjects	$n = 20$	$n = 168$
Source of subjects	VA hospital	Selected by E_A locally in university precinct
Procedure	Individual testing	Individual testing
Display	Tachistoscope	Tachistoscope
Exposure duration	Individually determined. Increased from 4 to 20 msec. (modal value) across 47 exposures	Individually determined + post-experimental check. Fixed duration throughout
Supraliminal material	Line-drawing of face	Line-drawings of eight faces
Response indicators	Verbal report of '*mood* changes', subsequently rated by two judges. Report latency. Imagery reports	'*Mood*' ratings on seven-point HAPPY–ANGRY scale
Design	HAPPY *v.* ANGRY: within-subjects variable. Same sequence of HAPPY and ANGRY for all subjects	HAPPY *v.* ANGRY: between-subjects variable
Experimenters	E_1 to control display. E_2, ignorant of stimulus, to prompt subjects' responses.	Seven assistant experimenters

Analysis of variance of the pre-experimental log thresholds indicated a small but significant main effect of experimenters ($F = 2{\cdot}40$; d.f. = 6, 140; $P < 0{\cdot}05$). There was no other significant main or interaction effect, although thresholds tended to be higher when E_As were not blind than when blind with respect to the cue ($F = 4{\cdot}36$; d.f. = 1, 6; $0{\cdot}10 > P > 0{\cdot}05$); this was true for six of the seven E_As. The reliability of this effect is emphasized by an analysis of variance of the post-experimental log thresholds; the bias factor was now significant ($F = 8{\cdot}74$; d.f. = 1, 6; $P < 0{\cdot}05$), and by this stage all seven E_As were recording higher thresholds in the not blind condition. However, the main effect due to experimenters was no longer significant in the post-experimental threshold session ($F = 0{\cdot}92$; d.f. = 6, 140; $P > 0{\cdot}05$).

Because almost twice as many males as females were selected by the E_As, an analysis of the ratings with sex as a factor had to sacrifice one of the independent variables. A subsidiary analysis of variance was computed ignoring the experimenter factor but retaining the cue and bias factors. This unequal-cell analysis of variance, by the unweighted means method, was based on the mean rating per subject, and revealed a significant sex × cue × bias interaction ($F = 7{\cdot}79$; d.f. = 1, 157; $P < 0{\cdot}01$). *Post hoc* comparisons (Scheffé, 1959) between the happy and angry conditions revealed a significant difference for male subjects tested by not blind E_As ($F = 7{\cdot}36$; d.f. = 1, 157; $P < 0{\cdot}01$). This difference (see Table 7) is the only indication that E_As were biasing their subjects directionally.

Preliminary discussion of Experiment III

The main analysis contained no support for the subliminal perception hypothesis and in this sense Expt. III failed to replicate its source study. It is important again to consider whether procedural differences were sufficient to account for the failure to replicate.

Inspection of Table 8 suggests that there was little difference in terms of sensitivity to subliminal perception effects. The difference between response indicators is a major one but it is not clear where the advantage in sensitivity lies. The rating-scale procedure was preferred on several grounds: it seemed to emphasize to the subject and E_A what were the appropriate dimensions of judgement; it seemed that experimental error would be reduced by excluding judges as a source of variability; and it supplied a more immediate and direct means of quantifying response processes. On the other hand, it could be countered that the use of verbal reports in the original study was more suitable because it involves less restriction of the influence of the subliminal cue. Unfortunately there is no convincing independent empirical justification for the use of one method as opposed to the other.

Sensitivity to the experimental treatment can also be discussed in relation to the design status of the cue factor. In the original study, the subject was presented with both HAPPY and ANGRY as subliminal cues, while in the replication the subject was presented with either HAPPY or ANGRY but not both. All else being equal, a within-subjects comparison is generally more sensitive to a given treatment effect than a between-subjects comparison, and so this seems to indicate a sensitivity advantage to the source study. But it is not obvious that all else was equal since the dependent variable and the statistical methods that were applied (including of course the associated degrees of freedom) also differed. It is therefore not easy to decide which study would be at an advantage regarding sensitivity to subliminal perception effects, particularly since theoretical arguments can be made in both cases that obscure the question further.

The failure to replicate could be attributed to differences in the subject populations. It has been suggested (Dixon, 1971) that individuals in whom primary process thinking is more dominant may be more receptive to subliminal stimulation. However it is not clear that there is an advantage in this respect one way or the other except insofar as the source study used hospital patients most of whom seemed to have had psychiatric symptoms.

The defences against artifacts also differ between the two studies. The exposure duration in the source study was progressively increased towards the pre-experimental threshold level and in this way it would seem to have given some opportunity for the use of partial cues. In the replication study greater protection against the use of partial cues was achieved by eliminating subjects whose post-experimental threshold fell to the level of the critical duration used during the experiment.

Experimenter bias is another potential source of artifacts. The blind *v.* not blind manipulation of the replication study failed to produce an effect on the main dependent variable, which suggests that experimenter bias is not the critical difference between the two studies. On the other hand, when subject sex was taken into account, the replication study did yield evidence of experimenter bias; positive results were

obtained with male subjects when the E_As were not blind and able to bias. This interaction effect was quite small and accounted for no more than 4 per cent of the total variance, and since the analysis was conducted *ex post facto*, the result has only a tenuous status.

Somekh & Wilding (1973) conducted a successful constructive replication of the source study of Expt. III, using a binocular rivalry paradigm to arrange that the subject was unaware of the cue. This was a technically impressive study which incorporated a device for 'hopefully eliminating any possibility of experimenter bias ...'; the 20 stimulus slides were numbered and presentation orders were subsequently decided by reference to random number tables. But this is not a method, however well intentioned, that prevents the experimenter from being informed about the stimuli. Furthermore, partial knowledge of the stimulus sequence could be enough to mediate bias. Indeed, this becomes increasingly plausible in Somekh & Wilding's study when it is realized that the stimuli, which were replicated, effectively formed only five distinct categories. Thus the incidental acquisition of the relevant associations was a real possibility. The basis for directional influence by the experimenter, irrespective of his intent, could reside in partial information about the critical stimulus categories (or category).

Although in the crucial subliminal session of Somekh & Wilding's study there was a significant difference in average mood ratings when the critical cues (HAPPY and SAD) were presented, the results in detail present a difficulty for the subliminal perception hypothesis which can be resolved by the partial information assumption. In Expt. I the SAD mean seems to have been homogeneous with the several control means but significantly different from the HAPPY mean. The situation in Expt. II changed somewhat, so that it was now the HAPPY mean which was similar to the control means but still significantly different from the SAD mean. It is conceivable that this pattern of results could have come about through the experimenter's inadvertent knowledge of one or two stimuli (HAPPY in Expt. I and SAD in Expt. II).

Smith *et al.* (1959) also tried to eliminate experimenter bias, with E_1 to operate the tachistoscope and E_2 to record and prompt the subject's responses. But since the presentation sequence was the same for all subjects and included blocks of HAPPY and ANGRY stimuli, it seems possible for E_2 to have acquired some knowledge of the stimulus order. The judged mood of the face varied according to the cue but the effect was more marked for HAPPY than for ANGRY. Thus the incomplete nature of the findings of both Smith *et al.* (1959) and Somekh & Wilding (1973) is consistent with the view that experimenter bias was not successfully ruled out and that it unwittingly became effective via the experimenter's partial knowledge of the stimulus sequence.

GENERAL DISCUSSION

Experimenter bias

From the results of these studies it is tempting to conclude that experimenter bias is not a sufficiently drastic or far-reaching phenomenon to be of further concern to subliminal perception theorists. This would be premature in view of the suggestion of positive bias in Expt. III, the discussion of Somekh & Wilding's (1973) study and Dixon's (1971) counter-claim about negative bias. What seems clear is that experi-

menter bias is, like subliminal perception, certainly a fragile phenomenon. Whether the experimenter bias effect is so specific in the case of subliminal perception as to be dependent on the subject's sex, as these results suggest, remains for a more searching study to determine. It also remains for further experiments to discover whether the experimenter's influence is stronger in less structured tasks, which may be more prone to experimenter bias effects (Masling, 1966; Barber & Silver, 1968).

The generality of experimenter bias has been questioned by several writers (Barber & Silver, 1968; Compton, 1970; Stewart, 1971). Discussions of the generality of the effects have emphasized the inconsistency of the phenomena, the unrepresentative nature of the experimenters employed, and the limited range of experimental tasks and situations that have been studied. In relation to the third criticism, it is striking that the work on experimenter effects rarely touches issues of any substance. There is much evidence of a psychologically and theoretically impoverished kind (e.g. person perception studies with no theoretical content). Studies like the present ones, where there is a general psychological hypothesis, are exceptions. In the light of the difficulties experienced here of pinpointing experimenter bias effects, one could speculate that bias may be more likely in circumstances where the experimenter has to make an 'effort after meaning'. Paradoxically therefore it may be the case that experimenters bias *less* where they do not have to infer or define for themselves the larger objectives of the study.

It is a simple matter for an investigator to attribute subliminal perception – or any other phenomenon he finds objectionable – to the operation of experimenter bias and demand characteristics. For this assertion to be taken seriously it is necessary to supply more than analogical evidence and this study has initiated this process for subliminal perception research. There are signs in the data presented above that suggest that experimenter bias may be effective in this context. The weakness of the effect is not reassuring, since the phenomena of subliminal perception are themselves reputedly small and difficult to detect. If further studies confirm these tentative conclusions it will be necessary to make explicit what the 'demand characteristics' are, how they manifest themselves, how they function, how and when experimenters bias their subjects. The present evidence that there were experimenter effects on threshold estimates indicates that experimenters can affect measures used in perception research. If the dependent variable in a study itself depends on the prior determination of a critical duration or intensity then it too could be subject to indirect but systematic influence.

Negative bias

In the light of the objectives of the present investigation it is difficult for us to enthuse about the possibility that the results were due to negative bias. Before discussing this issue it should be noted that there is a more compelling reason for being guarded about the use of negative bias as an explanatory concept. The difficulty arises because negative bias predicts no subliminal perception effect, an outcome which also follows if the subliminal perception hypothesis is not valid.

Moreover, a joint hypothesis consisting of the subliminal perception hypothesis (for positive instances of the effect) and negative bias (for null instances) seems *a priori*

to be incontrovertible, and therefore unacceptable in such a comprehensive form. It is necessary not only to state ground-rules for obtaining subliminal perception effects, but also for their eradication. Our preference, in line with the rationale for these experiments, would be for negative bias to be manipulated and demonstrated experimentally. This is not to say that it would be unprofitable to seek for common features of studies which fail to demonstrate subliminal perception, since clearly this is likely to provide the major source of ideas as to what constitutes negative bias.

It will perhaps be appreciated that we have little to offer in the way of pointing to potential defects of the present experiments in relation to negative bias. It has been suggested (by a consulting editor) that the presence of the experimenter, although entirely concealed, in the room with E_A and the subject during Expts. I and II may have served to introduce negative bias, although this could not be a criticism of Expt. III. On the other hand, common to all three experiments was the fact that all E_As were relatively inexperienced. Their management of the experimental procedures may have served to obscure a subliminal perception effect, although it is arguable whether this is a case of negative bias. It is of course easy to use this line of reasoning to suggest that it is only with experience that an investigator can come to produce positive results in subliminal perception studies. However, if the effect was one of E_As injecting 'noise' into the data and hence swamping the effects of subliminal perception, it should follow that unaccounted for variability would be high. Direct comparison of error variances between source and replication studies was not possible because non-parametric analyses were used originally. The indirect and perhaps hazardous approach was therefore adopted of computing the mean squares for common main and interaction effects in Expt. I based on Spence & Holland's (1962) mean recall scores. The mean squares for the source study for cue, recall category and cue × recall category were found to be 2·98, 19·05 and 7·56 respectively (unweighted means solution), and in the replication the corresponding values were 0·88, 53·78 and 5·34 (Table 1). These variances seem quite comparable and, in particular, the crucial interaction variance was larger in the source study and would also have been declared significant had it occurred here. This suggests that the present error variances were not inflated and that E_As were not introducing spurious variability into the data on a gross unselective basis. If negative bias was active, then it was of a somewhat subtle and selective kind. This is a little puzzling, since it seems to be at odds with how it might be expected to be manifest according to Dixon's (1971) discussion of the problem. It appears that this cannot be easily resolved without appropriate research. This seems to require an independent investigation which incorporates an unequivocal demonstration not only of negative bias but also of subliminal perception.

Replicating subliminal perception effects

The procedural differences between the source and replication studies were intended to favour the latter in terms of sensitivity to subliminal perception,* and in any event it seems reasonable to describe the latter as constructive replications (Lykken, 1968). Failure to replicate the original findings does not, of course, constitute disproof of the subliminal perception hypothesis, although it reduces the generality of the data base on which the hypothesis rests. It seems at least to follow that subliminal perception effects are weak and can easily be made to disappear. This weakness of the phenomenon (Dixon, 1971) is in fact one of the strengths of the subliminal perception hypothesis, since it is easy to argue for the inadequacy of any study that fails to obtain positive evidence.

In terms of statistical significance the present findings did not confirm the source experiments. It could be argued that the numerical pattern of differential recall scores in Expt. I was the same as Spence & Holland's, and the ratings in Expt. III were on average in the expected direction. But this approach would have to be applied across the board, and it breaks down for Expt. II. Furthermore, to achieve significance in Expt. I, twice as many E_As would have to be employed, and Expt. III could be continued indefinitely without the cue effect becoming statistically significant. Even in these terms it is hard to construe the present data as support for the subliminal perception hypothesis.

The present experiments seem to cast doubt on the replicability of the original findings and not necessarily because of the possibility of experimenter bias. Dixon (1971) has called attention to the need to replicate his own experiments. The need is evidently more extensive if the well-publicized effects of the present source studies can be so easily dissipated as they were here. It should be emphasized that the original findings may be valid and that the present outcome could be due to negative bias. Otherwise the most parsimonious explanation of our failure to replicate is that the two source studies involved Type I errors and we were therefore chasing a statistical will-o'-the-wisp.

This research was supported by a grant from the Social Science Research Council. A version of this paper was presented to the Annual Conference of the British Psychological Society, Nottingham, 1972.

* A reviewer has suggested that the source study for Expt. I was more sensitive to subliminal perception effects than the replication, since the exposure duration in the former was lower than in the latter. If subliminal perception effects are inversely related to exposure duration then this could explain why the source study obtained positive results and the replication did not. But because there were instrumentation differences as well as unmentioned factors like luminance, dark adaptation and so on, it would be hazardous to compare sensitivities on the basis of exposure duration alone. Moreover, it seems likely that the exposure duration used by Spence & Holland was quite close to threshold, since they found it advisable to discard some data because of doubts about the subliminality of the stimulus. This suggests that the exposure durations were effectively very similar.

REFERENCES

BARBER, T. X. & SILVER, M. J. (1968). Fact, fiction, and the experimenter bias effect. *Psychol. Bull.* **70** (6, pt 2), 1–29.

COMPTON, J. W. (1970). Experimenter bias: reaction time and types of expectancy information. *Percept. mot. Skills* **31**, 159–168.

DEESE, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.* **5**, 305–312.

DIXON, N. F. (1971). *Subliminal Perception: the Nature of a Controversy.* London: McGraw-Hill.

HABER, R. N. & HERSHENSON, M. (1965). The effects of repeated brief exposures on the growth of a percept. *J. exp. Psychol.* **69**, 40–46.

HASLERUD, G. M. (1964). Perception of words as a function of delays between and summation of subliminal exposures. *Percept. mot. Skills* **19**, 130.

KIRK, R. E. (1968). *Experimental Design: Procedures for the Behavioural Sciences.* Belmont, Calif.: Brooks/Cole.

LYKKEN, D. T. (1968). Statistical significance in psychological research. *Psychol. Bull.* **70**, 151–159.

MASLING, J. (1966). Role-related behaviour of the subject and psychologist and its effects upon psychological data. In D. Levine (ed.), *Nebraska Symposium on Motivation.* Lincoln: University of Nebraska Press.

NEISSER, U. (1967). *Cognitive Psychology.* New York: Appleton-Century-Crofts.

ROSENTHAL, R. (1966). *Experimenter Effects in Behavioural Research.* New York: Appleton-Century-Crofts.

ROSENTHAL, R. (1969). Interpersonal expectations: effects of the experimenter's hypothesis. In R. Rosenthal & R. L. Rosnow (eds.), *Artifact in Behavioural Research.* New York: Academic Press.

ROTHKOPF, E. Z. & COKE, E. U. (1961). The prediction of free recall from word association measures. *J. exp. Psychol.* **62**, 433–438.

SCHEFFÉ, H. (1959). *The Analysis of Variance.* New York: Wiley.

SMITH, G. J. W., SPENCE, D. P. & KLEIN, G. S. (1959). Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.* **59**, 167–177.

SOMEKH, D. E. & WILDING, J. M. (1973). Perception without awareness in a dichoptic viewing situation. *Br. J. Psychol.* **64**, 339–349.

SPENCE, D. P. & HOLLAND, B. (1962). The restricting effects of awareness: a paradox and an explanation. *J. abnorm. soc. Psychol.* **64**, 163–174.

STEWART, C. G. (1971). Consistency, generality, magnitude, and significance of experimenter expectancy effects in human research. *Psychol. Rec.* **21**, 449–458.

(*Manuscript received* 30 *May* 1973; *revised manuscript received* 14 *October* 1974)