

Investigations of perception and imagery using CAEs: The role of experimental design and psychophysical method

JACK BROERSE

Brisbane College of Advanced Education, Mount Gravatt, Queensland, Australia

and

BORIS CRASSINI

University of Queensland, St. Lucia, Queensland, Australia

Orientation-contingent color aftereffects (CAEs) were measured using 174 observers randomly allocated to one of eight experimental conditions formed by the factorial combination of three two-level factors. These were experimental design (pretest and posttest measures taken vs. posttest measures taken alone), psychophysical method (forced-choice vs. magnitude estimation), and inspection procedure (inspection of real contours vs. inspection of imagined contours). In addition to color responses, written introspective reports of the strategies used in providing color responses were obtained. As a basis for the rejection of observers using inappropriate strategies, reported strategies were coded as appropriate or inappropriate. An analysis of the color reports of all observers was performed, as well as two analyses of the color reports of only those observers identified as using appropriate response strategies. In all cases, the results showed that CAEs were reported after inspection of real contours irrespective of the experimental design or psychophysical method used. There were no significant CAEs reported in any conditions that involved the imagination of contours. A separate analysis of the relationship between observers' response strategies, the coding of these strategies, and observers' actual color reports revealed some problems concerning the use of such introspective techniques. The implications of these findings are discussed in terms of theoretical assumptions about the nature of imagery and perception.

A criticism of experimental investigations of the relationship between perception and imagery is that observers' responses may be influenced by the demand characteristics of the situation and/or the observers' knowledge of, or expectations about, the phenomenon being studied (Intons-Peterson & White, 1981; Mitchell & Richman, 1980; Richman, Mitchell, & Reznick, 1979a, 1979b). One perceptual phenomenon that is, to some extent, less open to these criticisms is the McCollough (1965) effect, a spatially contingent color aftereffect (CAE). CAEs are generated by prolonged inspection of particular combinations of color and contour orientation (e.g., red vertical contours together with green horizontal contours). The CAEs are manifested using achromatic test stimuli that are typically seen by observers as colored (in this example, the vertical contours appear green and the horizontal contours, red). That is,

CAEs are negative and, in this respect, are counter to any association between, or expectation about, color and contour orientation that may have been built up during inspection. The negative nature of CAEs provides an informal, but reliable, check that observers are responding on the basis of the appearance of the test stimuli, rather than on the basis of what they can remember about the pairings of color and contour orientation that occurred during inspection.

Within these terms, the claim that orientation-specific color effects could be generated by requiring subjects to imagine vertical (horizontal) contours on physically presented patches of red (green) color (Finke & Schmidt, 1977, 1978) may be defended as providing "perhaps the strongest answer so far to criticisms that subjects always use their knowledge about objects or expectations for performance to generate predicted responses in mental imagery tasks" (Finke, 1980, p. 121; see also Kosslyn, 1981; Kosslyn, Pinker, Smith, & Schwartz, 1979). While it may be true that imagery-induced color aftereffects (IICAEs) are less open to such criticisms, they are not free from other criticisms based on discrepancies between

Direct reprint requests to Jack Broerse, Department of Psychology and Special Education, Brisbane College of Advanced Education, P.O. Box 82, Mount Gravatt, Queensland 4122, Australia, or to Boris Crassini, Department of Psychology, University of Queensland, St. Lucia, Queensland 4067, Australia.

the observable properties of IICAEs and CAEs. For example, differences between CAEs and IICAEs have been reported for (1) the relative strength of CAEs [IICAEs are usually described as being significantly weaker (less saturated?) than CAEs]; (2) the nature of the colors reported (CAEs are invariably negative, whereas IICAEs can be either positive or negative); and (3) the dependence of CAEs and IICAEs on the self-reported imagery ability of observers (CAEs are independent of self-reported imagery ability, whereas IICAEs are reported as being dependent on this ability).

We have discussed these and other issues more fully elsewhere (Broerse & Crassini, 1980, 1981), and have indicated that perhaps the most important difference between CAEs and IICAEs is the ease with which they can be generated. CAEs are robust, compelling, and easily produced phenomena, whereas IICAEs are elusive phenomena requiring special procedures for their production (see below). Indeed, in the second of two experiments (Broerse & Crassini, 1980), we were unable to produce IICAEs at all. In that experiment, we used a pretest/posttest experimental design with a variation of magnitude estimation as our psychophysical measurement technique. It has been suggested that IICAEs "require rather sensitive testing procedures for their resolution" (Finke & Schmidt, 1977, p. 606), and it may be that a pretest/posttest design plus magnitude estimation is an insufficiently "sensitive" means to "resolve" IICAEs (Finke, 1981; Kunen & May, 1981). This question was addressed in the experiment to be reported by comparing directly a pretest/posttest design and magnitude estimation procedure with the posttest-alone and forced-choice procedure used by Finke and Schmidt (1977, 1978).

In addition to this question, one crucial aspect of the "resolution" of IICAEs was investigated. The "resolution" of IICAEs as performed by Finke and Schmidt (1977, 1978) involved two components: the first was the measurement of an observer's color experience, and the second was the successful elimination of those observers who used what Finke and Schmidt (1977) termed inappropriate response strategies. The most frequently occurring inappropriate response strategies were those based on recall of the color-contour orientation relationship used during the inspection phase of the experiment (i.e., association strategies). These were identified by analysis of written introspective post hoc reports made by observers. The importance of this part of the resolution of IICAEs was stressed by Finke (1981), who stated that "there is a strong tendency in this type of experiment for subjects to think that they are supposed to give as responses the same colors, in association with pattern orientation, that had been presented (or imagined) during adaptation procedures" (p. 94). It should be pointed out that the occurrence of IICAEs

in Finke and Schmidt (1977) was not different from chance unless observers using such inappropriate strategies were eliminated from the sample. Furthermore, it was reported that 24 of 49 observers in the imagery condition were classified as using inappropriate response strategies, while only two out of 45 observers exposed to the physically present colors and contours were so classified. While these data support the claim that there is a strong tendency for observers to report inappropriate strategies in imagery conditions, they also suggest that this does not occur when real contours are used during inspection. More importantly, since Finke and Schmidt (1977) do not compare the two, there is no direct evidence to suggest that the strong tendency to report inappropriate strategies is reflected in observers' *actual color responses*. It is clear, therefore, that the "sensitivity" of any methodology used to investigate IICAEs must be assessed not only in terms of measurement of an observer's experience of color but also in terms of how well particular types of color responses can be predicted by observers' introspections about response strategies. In the present investigation, introspective reports were used to eliminate observers who reported using inappropriate response strategies (in the manner recommended by Finke, 1981; Finke & Schmidt, 1977, 1978). In addition, the relationship between introspective reports and corresponding color reports was also analyzed.

Overview of Design, Data Collection, and Analyses

A three-way, between-subjects factorial design was used so that individual cells and combinations of cells represented replications and extensions of the experiments of Broerse and Crassini (1980) and Finke and Schmidt (1977).

The three factors were:

Experimental design: Two levels—pretest/posttest (PPT) vs. posttest alone (PT).

Psychophysical method: Two levels—forced-choice (FC) vs. magnitude estimation (ME).

Type of inspection procedure: Two levels—real contours vs. imagined contours.

The design and procedure of Finke and Schmidt (1977) were replicated in Conditions 5 and 6, and those of Broerse and Crassini (1980) in Conditions 3 and 4 (see Table 1). The remaining conditions represented logical extensions of these designs and procedures in a factorial context. Measurement of observers' color experience was carried out by procedures described below. In addition, each observer was required to provide introspective verbal reports about his/her color responses. The verbal self-reports were judged as to whether the observers had used association or non-association response strategies. Ten naive coders using the criteria outlined by Finke and Schmidt (1977)

performed this task. Analyses of color responses are reported for all observers and for only those observers judged not to be using association response strategies. The introspective reports (i.e., how observers said they responded) were also analyzed to determine (1) their relationship to the direction of color reports (i.e., how observers actually did respond), and (2) whether the nature of the report (i.e., association vs. nonassociation) was influenced by the type of inspection procedure used.

METHOD

Subjects

One hundred and seventy-four students enrolled in undergraduate courses at the Brisbane College of Advanced Education participated. All participants were naive as to the purpose of the experiment, and none reported having previously participated in perception or imagery experiments. In addition, 10 coders were used to judge observers' introspective reports. Of the 10 coders, 3 were selected randomly from among the participating observers, 3 were volunteers drawn from the student population (they had not participated in the experiment), and 4 were volunteers drawn from the teaching faculty in the Department of Psychology and Special Education.

Materials and Apparatus

Inspection and test stimuli were projected onto a white screen using a Kodak Carousel projector fitted with a 150-W quartz-iodide filament. All stimuli were circular with an 85-cm diameter, and gratings (when used) had a square-wave luminance profile. The spatial frequency of the gratings was 1.7 cycles/deg at a viewing distance of 3 m and 2.8 cycles/deg at 5 m. Inspection gratings consisted of only horizontal (H) or vertical (V) contours, and test stimuli consisted of V and H segments in the top, bottom, left, or right halves of the display (i.e., V/H, H/V, H-V, V-H). In all inspection conditions involving the physical presentation of color, red was produced by using Kodak Wratten Filter No. 26 and green by using Wratten filter No. 55. In both the real and imagery conditions, red was always combined with vertical contours and green with horizontal ones.

Procedure

Observers were tested in groups ranging in size from 16 to 29, with individuals seated as close as 3 m and far as 5 m from the screen on which stimuli were projected. The groups were allocated randomly to one of the eight conditions shown in Table 1. The sequence of the procedures in which they took part was: (1) pretest

color reports (Conditions 1 to 4), (2) inspection (of real or imagined contours), (3) posttest color reports, and (4) written introspective reports.

Pretest color reports. Observers were told that they would be participating in an experiment in which they would be requested to make judgments concerned with the perception of color. They were asked to postpone any questions about the nature of the experiment until a final debriefing session. The observers were then told the psychophysical measurement technique (FC or ME) appropriate to their experimental condition. The four test slides were presented twice in random order, each slide appearing for approximately 5 sec, and the observers recorded their color judgments on a score sheet. The observers in the FC conditions received instructions taken directly from Finke and Schmidt (1977); those in the ME conditions received instructions as per Broerse and Crassini (1980; see Appendix). Score sheets for all observers were contained in booklets in which each sheet was separated by a blank page. All observers were required to complete the response blanks without referring to previous responses.

Inspection. The observers in the real inspection conditions were shown red vertical (10 sec) and green horizontal (10 sec) slides in alternation for 5 minutes. In the imagery inspection conditions, the observers viewed red (10 sec) and green (10 sec) homogeneous color disks in alternation, for the same length of time, and were required to imagine that the red disks contained vertical bars and the green disks horizontal bars. As an aid to imagination, the observers were shown achromatic gratings of vertical and horizontal bars identical in size and spatial frequency to those used in the real condition. The instructions to imagine bars were taken (with slight modification) from Finke and Schmidt (1977, see p. 601) and are presented verbatim in the Appendix. The observers were requested to maintain steady gaze and an upright head position throughout the inspection phase.

Posttest color reports. The observers in the PPT conditions were reminded of the procedure (either FC or ME) for responding to the test stimuli. The observers in the PT conditions were briefed on the techniques for making color responses, as outlined above. Care was taken to ensure that both PPT and PT groups underwent a similar delay between cessation of the inspection phase and beginning of the test phase. The four test slides were then exposed twice in random sequence.

Introspective reports. At the completion of the posttest color report phase, all observers were requested to write an introspective report using the procedures of Finke and Schmidt (1977).

Coding of Introspective Reports

A perusal by us of the introspective reports revealed that some contained cues as to the experimental condition in which the data were collected. For this reason, we departed from the "bias control" procedures adopted by Finke and Schmidt (1977) to classify

Table 1
Overall Factorial Design of the Experiment With Total Ns for Each Condition

Psychophysical Method	Experimental Design							
	Pretest, Posttest				Posttest			
	Forced-Choice		Magnitude Estimation		Forced-Choice		Magnitude Estimation	
	Imagery	Real	Imagery	Real	Imagery	Real	Imagery	Real
Inspection Procedure	Imagery	Real	Imagery	Real	Imagery	Real	Imagery	Real
Condition	1	2	3	4	5	6	7	8
Total N	24	26	23	17	29	19	20	16
Nonassociation								
N (Unanimous)	8(33%)	15(57%)	9(39%)	8(47%)	10(35%)	12(63%)	7(35%)	10(63%)
Nonassociation								
N (Majority)	16(67%)	21(81%)	21(91%)	17(100%)	17(59%)	16(84%)	13(65%)	11(69%)

Note—Also shown for each condition are the Ns classified as nonassociation by written verbal report (see text).

observers' response strategies. Those authors coded the reports themselves (before referring to the actual color report data). Instead, we decided to use coders who were unaware of the particular issues involved in the experiment, and, except for the three observers who participated in the experiment, were completely naive about the nature of CAEs. The pages of individual response booklets were numerically identified and the introspective reports were separated from the remaining data of each individual. This ensured that the coding of introspective reports was not influenced by the color responses.

The coders were informed that an experiment had taken place which required observers to make color judgments about the color appearance of some test slides. They were also told that the task was sometimes difficult, and that some observers may have used criteria other than the actual appearance of the test slide to make their color responses. The nature of these criteria (e.g., an association strategy) was explained in terms of the "bias control" and Appendix sections of Finke and Schmidt (1977). Examples of "relevant" and "irrelevant" strategies reported by Finke and Schmidt (1977) (e.g., part of a test slide appeared "more salient") were used as illustrations. No numerical information (e.g., the numbers of subjects classified as "association" subjects) was provided. In addition, the actual test slides were shown as an example of the task required of observers in the experiment. The coder's task, therefore, was to read each verbal report and decide whether it indicated that observers had based their color reports on the appearance of the test slides (as per experimental instructions) or whether their color reports were based on some inappropriate strategy.

RESULTS AND DISCUSSION

Analysis 1: Color Responses, All Observers

For observers in FC conditions, color responses were converted to color scores using the methods of Finke and Schmidt (1977). When observers reported the half of the test slide containing horizontal contours as appearing more red (less green) than the half containing vertical contours, they received a color score of +1. (Note that this represents the color-contour relationship typically found in CAE experiments, that is, a nonassociation color response). A color score of 0 was given to responses in which vertical contours were said to be more red (less green) than vertical contours. (Note that this represents the color-contour relationship used during inspection; that is, it may be described as an association color response.) Each observer's maximum possible color

score was +8 (all test slides responded to appropriately), and a score of +4 represented chance responding (equal numbers of association and nonassociation color responses). Pretest color responses were scored using the same convention, even though the inspection phase had not occurred.

In ME conditions, color responses were scored in the manner previously reported by Broerse and Crassini (1980). All judgments were standardized by dividing an observer's separate magnitude estimates for both halves of the eight test slides by the maximum estimate reported by that observer. Nonassociation color responses were scored as positive, and association color responses were scored as negative. The maximum possible color score was thus +16 (all color responses of the same magnitude, and nonassociation in nature) and the minimum score -16 (all color responses of the same magnitude, and association in nature). A color score of 0 would occur if observers reported no color, or if their color responses were random. As for the FC conditions, the same scoring convention was used for both pretest and posttest responses.

The mean pretest and posttest color scores for the experimental groups tested with the FC procedure are presented in Table 2; those for the groups tested with the ME procedure are presented in Table 3. A series of *t* tests was carried out to determine whether each mean was different from that expected by chance for the particular group. Since it was expected that posttest color scores would be greater than chance, one-tailed tests were applied; for pretest scores, two-tailed tests were used. Tables 2 and 3 contain the results of these *t* tests, which show that all pretest color scores, whether obtained via FC (Conditions 1 and 2, Table 2) or ME procedures (Conditions 3 and 4, Table 3), were not different from chance. Following inspection of real contours, the observers reported color responses of the typical CAE type. These were significantly different from chance irrespective of the psychophysical method used (Conditions 2 and 6, Table 2; Conditions 4 and 8, Table 3). However, no significant effects were found in any of the conditions in which observers were required to imagine

Table 2
Mean Color Response Scores (Together With *t* tests for Differences From 4.0) for All Observers in Forced-Choice Conditions

	Condition							
	1		2		5		6	
	Imagery		Real		Imagery		Real	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Chance	4.0	4.0	4.0	4.0		4.0		4.0
Obtained	4.7	4.3	3.2	7.9		3.1		7.5
<i>t</i>	1.06	0.45	-1.35	25.67		-1.62		10.00
Significance	n.s.	n.s.	n.s.	<i>p</i> < .001		n.s.		<i>p</i> < .001

Note—Two-tailed *t* tests were applied for both pretest scores. See Table 1 for *N*s in each condition.

Table 3
Mean Color Response Scores (Together With *t* Tests for Differences From 0.0) for All Observers in Magnitude-Estimation Conditions

	Condition							
	3		4		7		8	
	Imagery		Real		Imagery		Real	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Chance	0.0	0.0	0.0	0.0		0.0		0.0
Obtained	-0.4	0.5	-0.1	11.1		-0.7		11.4
<i>t</i>	-1.17	0.86	-0.07	11.46		-0.48		13.54
Significance	n.s.	n.s.	n.s.	<i>p</i> < .001		n.s.		<i>p</i> < .001

Note—Two-tailed *t* tests were applied for both pretest scores. See Table 1 for *N*s in each condition.

contours during the inspection phase (Conditions 1 and 5, Table 2; Conditions 3 and 7, Table 3).

The data presented in Tables 2 and 3 are replications of Broerse and Crassini (1980) and Finke and Schmidt (1977) when all observers are included. Observers in Conditions 3 and 4 (which replicate Broerse & Crassini, 1980) and Conditions 5 and 6 (which replicate Finke & Schmidt, 1977) reported significant numbers of appropriate color responses following inspection of real contours, but responded at chance levels following inspection of imaginary contours. Inspection of the results of Conditions 1, 2, 7, and 8 shows the same pattern, so that when the data of all observers are considered, the experimental design and the psychophysical method used are irrelevant: Inspection of real contours generates CAEs, inspection of imaginary contours generates chance color responding.

Analysis 2: Color Responses, Nonassociation Observers

We pointed out in the introduction that an integral part of "resolving" IICAEs is the identification and removal of the data of those observers using inappropriate association strategies as a basis for their responses. We used 10 independent coders to classify observers' introspective verbal reports as association (A) or nonassociation (NA) in nature. For the purposes of analysis, each coder's classification was

treated as a dichotomous item (*A* = 1, *NA* = 0), making up a 10-item introspective report scale. An observer obtained an introspective report score of 0 when all 10 coders agreed that the verbal report indicated that the observer's color responses were made on the basis of a nonassociation strategy (i.e., on the basis of the appearance of the test slides). The overall interitem (i.e., intercoder) reliability coefficient (Cronbach's, 1970, α) was 0.91, and the point-biserial item-total correlations ranged from $r_{pbi} = 0.58$ to $r_{pbi} = 0.85$ (mean = 0.74, *SD* = 0.09). Both indices show a high degree of consistency between coders in interpreting and applying the criteria for classification of verbal reports as association or nonassociation.

Observers with an introspective report score of 1 or more were removed from the sample so that the remaining observers were judged unanimously to have used a nonassociation strategy in giving color responses. The mean pretest and posttest color scores (together with *t* tests from chance responding) for the nonassociation observers tested with the FC procedure are shown in Table 4; those tested with the ME procedure are shown in Table 5.

The pattern of ME results for nonassociation observers (Table 5) was essentially the same as that found for the total sample (Table 3). Pretest color scores and posttest color scores following imagination were not significantly different from those expected by chance. Posttest color scores following in-

Table 4
Mean Color Response Scores (Together With *t* tests for Differences From 4.0) for Observers Judged as Nonassociation Respondents in the Forced-Choice Conditions

	Condition							
	1		2		5		6	
	Imagery		Real		Imagery		Real	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Chance	4.0	4.0	4.0	4.0		4.0		4.0
Obtained	6.8	6.4	2.7	8.0		4.3		7.3
<i>t</i>	2.76	2.40	-1.43	40.00		0.29		6.02
Significance	<i>p</i> < .05	<i>p</i> < .05	n.s.	<i>p</i> < .001		n.s.		<i>p</i> < .001

Note—Two-tailed *t* tests were applied for both pretest scores. See Table 1 for *N*s in each condition.

Table 5
Mean Color Response Scores (Together With *t* tests for Differences From 0.0) for Observers Judged
as Nonassociation Respondents in the Magnitude-Estimation Conditions

	Condition							
	3		4		7		8	
	Imagery		Real		Imagery		Real	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Chance	0.0	0.0	0.0	0.0		0.0		0.0
Obtained	-0.1	-0.6	1.1	10.0		1.0		10.2
<i>t</i>	-0.10	-1.01	1.12	8.39		1.08		9.17
Significance	n.s.	n.s.	n.s.	<i>p</i> < .001		n.s.		<i>p</i> < .001

Note—Two-tailed *t* tests were applied for both pretest scores. See Table 1 for *N*s in each condition.

spection of real contours were significantly different from chance and in the typical CAE direction. Comparison of the results shown in Tables 2 and 4 reveals a change in the pattern of FC results, but only for imagery conditions. In Condition 1, the mean posttest color score was significantly different from chance and in the appropriate direction (i.e., ICAEs were reported). Furthermore, in Condition 5, removal of the data of association respondents produced a change in the mean posttest color score from 3.1 (all observers) to 4.3 (nonassociation observers). While not significantly different from chance responding, this at least represents a trend in the direction of ICAEs. However, this evidence in support of ICAEs is illusory: The significant posttest color scores in Condition 1 following inspection of imaginary contours is accompanied by a pretest color score also significantly different from chance in the same direction (see Table 5). When pretest and posttest scores are subtracted, the mean net effect is 0.4 (not different from zero).

In brief, the removal of the data of observers judged on the basis of their verbal reports to have used an association strategy resulted in little evidence for the occurrence of ICAEs. The culling procedure produced minimal changes in results from conditions in which inspection of real contours took place and from conditions in which imagery inspection and magnitude estimation were combined. In Condition 5, which replicated Finke and Schmidt's (1977) procedures, there was a hint of ICAEs, but this was not significant. Statistically significant ICAEs were found in the posttest phase of Condition 1, but were nullified by pretest color scores in the same direction and of the same magnitude.

Relationship Between Color Responses and Introspective Reports

The rationale underlying the identification and removal of observers using association strategies was to ensure that observers were responding on the basis of the appearance of the test slides. It is also possible

to classify observers on the basis of their color responses. Those who report vertical contours as red and horizontal as green in the present study may be described as association color respondents, while those reporting vertical and horizontal contours as green and red, respectively, are nonassociation color respondents. Observers giving no color responses or equal numbers of association and nonassociation color responses (i.e., with a forced-choice color score of 4.0, or a magnitude-estimation color score of 0.0) may be termed chance color respondents. We elaborate on this in our concluding remarks.

Table 6 shows the results of this classification for the overall number of observers (in imagery and real conditions) who were included in Analysis 1. The pattern of results reported in the original analysis (i.e., significant effects in real but not imagery conditions) is reflected by the number of observers in each of the three color response categories (i.e., association, nonassociation, and chance). For imagery conditions, the number of observers in each category was approximately equal, whereas, for real conditions, only one observer reported no color (i.e., chance) and one observer reported inappropriate association colors. Table 6 also shows the results of this classification procedure when applied to observers whose introspective reports were interpreted as reflecting nonassociation strategies by all 10 coders. These observers constituted the sample for Analysis 2 in which only 8 of 37 association color respondents were inappropriately included. While it appears that this unanimous criterion may have been particularly effective in eliminating association respondents from the sample, almost half the number of appropriate (i.e., nonassociation) color respondents were also eliminated. Indeed, in imagery conditions, the number of nonassociation respondents retained was less than half (i.e., 14 of 33; see Table 6). In overall terms, 45% of the original sample was retained compared to 72% in the original Finke and Schmidt (1977) study. Since the occurrence of ICAEs is as much dependent on the retention of nonassociation color respondents

Table 6
The Number of Association, Nonassociation and Chance (No Color) Respondents
Included in the Original and Subsequent Analyses

Color Respondents								
Association			Nonassociation			Chance		
Imagery	Real	Total	Imagery	Real	Total	Imagery	Real	Total
Analysis One: All Observers (N = 174)								
36	1	37	33	76	109	27	1	28
Analysis Two: Unanimous Criterion* (N = 79)								
7	1	8	14	44	58	13	0	13
Analysis Three: Majority Criterion** (N = 132)								
18	1	19	26	63	89	23	1	24

*Introspective reports coded as nonassociation by 10 out of 10 coders (see text). **Introspective reports coded as nonassociation by 7 out of 10 coders (see text).

as it is on the elimination of association color respondents, any procedure that eliminates the former will also reduce the likelihood of obtaining significant ICAEs. The unanimous criterion used for introspective reports in Analysis 1, therefore, may have been too stringent. For this reason, an additional analysis was performed in which a greater proportion of nonassociation color respondents was retained. In this analysis (Analysis 3), all observers whose introspective reports were classified as nonassociation by 7 of 10 coders (i.e., a majority criterion) were retained in the sample (see Table 6). Using this criterion, 76% of the original sample was retained, a proportion that resembled more closely that reported by Finke and Schmidt (1977).

Color scores for nonassociation observers defined in terms of the majority criterion were analyzed as previously and are shown in Tables 7 and 8. The pattern of results in Tables 7 and 8 is essentially the same as that reported in Tables 2 and 3 for the total sample and in Tables 4 and 5 for the nonassociation respondents classified as using the unanimous criterion. Significant color scores (CAEs) were found (irrespective of the experimental design or psychophysical procedure)

following inspection of real contours. Significant posttest color scores (ICAEs) were found in one condition following imagination inspection, but these were nullified by pretest color scores of the same magnitude and direction (see Concluding Remarks). It is clear that the use of a majority criterion made little difference to the overall pattern of results obtained with previous analyses. The reason for this is evident in Table 6 in that while the majority criterion increased the proportion of nonassociation (i.e., appropriate) color respondents to 89/109, there were corresponding increases in the proportion of association (19/37) and chance (24/28) respondents. For imagery conditions in particular, the effects of gaining 12 appropriate color respondents were canceled by corresponding gains of 11 inappropriate (association) color respondents and 10 chance (no color) respondents. Paradoxically, because almost all chance and association color respondents were found in imagery conditions, the less stringent majority criterion works against the detection of significant ICAEs.

One aspect of the rationale underlying the use of introspective reports warrants further elaboration. On first impressions, the use of such reports to iden-

Table 7
Mean Color Response Scores (Together With *t* tests for Differences From 4.0) in the Forced-Choice Conditions
for Observers Classified as Nonassociation by a Majority of Coders of Verbal Reports

	Condition							
	1		2		5		6	
	Imagery		Real		Imagery		Real	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Chance	4.0	4.0	4.0	4.0		4.0		4.0
Obtained	5.6	5.6	3.1	7.8		3.6		7.4
<i>t</i>	2.09	1.90	-1.25	20.05		-0.57		8.39
Significance	n.s.	<i>p</i> < .05	n.s.	<i>p</i> < .001		n.s.		<i>p</i> < .001

Note—Two-tailed *t* tests were applied for both pretest scores. See Table 1 for *N*s in each condition.

Table 8
Mean Color Response Scores (Together With *t* tests for Differences From 0.0) in the Magnitude-Estimation Conditions for Observers Classified as Nonassociation by a Majority of Coders of Verbal Reports

	Condition							
	3		4		7		8	
	Imagery		Real		Imagery		Real	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Chance	0.0	0.0	0.0	0.0		0.0		0.0
Obtained	0.3	0.5	-0.1	11.1		0.4		10.7
<i>t</i>	0.73	0.82	-0.07	11.46		0.34		9.40
Significance	n.s.	n.s.	n.s.	$p < .001$		n.s.		$p < .001$

Note—Two-tailed *t* tests were applied for both pretest scores. See Table 1 for *N*s in each condition.

tify association color respondents appears to be highly effective, particularly when the unanimous criterion is used. Implicit in the use of this procedure is the assumption that by eliminating as many association respondents as possible, the number of nonassociation (i.e., appropriate) color respondents will be maximized. To achieve this, the introspective procedure must predict *both* types of response class (i.e., association and nonassociation) equally well. Yet the data shown in Table 6 clearly show this not to be the case. Using the unanimous criterion, for example, 29 of the 37 association color respondents and 58 of the 109 nonassociation color respondents were correctly identified. This difference in correct and incorrect identifications for association and nonassociation respondents was significant ($\chi^2 = 7.3$, $p < .01$). Using the majority criterion, 18 of the 37 association color respondents and 89 of the 109 nonassociation color respondents were correctly identified. The accuracy of classification remained significantly different for the two groups ($\chi^2 = 13.3$, $p < .01$), although in the opposite direction to that found when using the unanimous criterion.

Although association color respondents were found to occur almost exclusively in imagery conditions (Table 6), the differences between how well introspective reports identified association and nonassociation color respondents were not simply a function of whether observers were in imagery or real conditions. For nonassociation color respondents, there were no differences in the proportions of correct predictions between observers in imagery conditions and observers in real conditions (unanimous criterion, $\chi^2 = 2.46$, $p > .05$; majority criterion, $\chi^2 = 0.11$, $p > .05$). This comparison was not made for observers giving association color responses, since only one such response was obtained in conditions in which real contours were inspected.

These results indicate that the degree of accuracy with which observers were correctly eliminated (on the basis of their introspective reports) varied as a function of whether the color responses given were of an association or nonassociation type together with the criterion (unanimous vs. majority) adopted to

classify respondents. While the second of these factors was due to our use of multiple coders, even the use of a single coder, or a forced consensus between more than one coder, would not have resolved the underlying problems arising out of the relationship between what observers do (i.e., in terms of their color responses), what observers report doing (i.e., in terms of their introspective reports), and how such introspective reports are interpreted (see below).

CONCLUDING REMARKS

In the present study, we attempted to replicate and extend systematically the procedures and results previously reported by Broerse and Crassini (1980) and Finke and Schmidt (1977). Our replication of both studies was successful to the extent that we were able to generate strong and convincing CAEs following inspection of real (as opposed to imagined) contours. The effects were equally convincing regardless of (1) whether observers who had reported using "association" strategies were retained in or excluded from the sample, (2) whether the design involved both pretest and posttest measures or posttest measures alone, and (3) whether the measurement procedures were based on FC or ME techniques. Following inspection of imagined contours, we were able to replicate the results of both studies: When all observers were included, IICAEs were no better than chance. Our replication of the most crucial part of the Finke and Schmidt (1977) study, however, was not successful. When observers were eliminated on the basis of their introspective reports, there was still no evidence of IICAEs.

The experimental conditions represented the factorial combination of the two factors suggested by Finke (1981) as underlying previous failures to demonstrate IICAEs: These were the differences in sensitivity associated with particular design and measurement techniques. Since chance levels of responding occurred with IICAEs across all conditions involving imagery, there was no support for these conjectures. However, the pattern of scores obtained with obser-

vers in Condition 1 (e.g., see Table 4) suggests an explanation of the effects previously reported by Finke and Schmidt (1977). The pretest scores of nonassociation observers in Condition 1 were significantly greater than chance. That is, these observers reported "CAEs" in the absence of color/contour orientation (either real or imagined), a phenomenon similar to that previously reported by Stanley and Hoffman (1976). The fact that such color responses occur inconsistently indicates that a pretest/posttest design is mandatory in studies involving CAEs. If no pretest scores are available, posttest scores cannot be unequivocally attributed to the experimental manipulation: They may simply reflect some predisposition of the observer. Thus, when the pretest and posttest scores of observers in Condition 1 are taken in conjunction, it is clear that instructions to imagine contours produced no changes in color responding.

A further issue raised by our data relates to the use of introspections about covert response strategies in studies of perception. Such techniques raise a special set of problems concerning, first, the nature of direct awareness of one's mental functions (see Natsoulas, 1970, 1978) and, second, the interpretation, by independent observers, of reports presumed to reflect such direct awareness. In the present study, an indication of these problems is revealed in the poor relationship found between how some observers said they responded (i.e., in terms of their introspective verbal reports), how they were judged to have responded, and how these observers actually did respond (i.e., in terms of their color reports). Although information of this nature is essential in an experimental context where observers are eliminated on the basis of their introspective reports, the problems concerning the correspondence between such reports and overt behavior remain. As Lieberman (1979) has pointed out, introspective reports are of interest only when their content can be corroborated by evidence from other sources, or when the same reports can be obtained with repeated observations in similar circumstances.

We should like to point out that our rejection of the notion that IICAEs constitute reliable phenomena is based on empirical considerations. Our own theoretical biases concerning the nature of perception and imagery are consistent with the expectation that interactions between perception and imagery should occur, even if only in the trivial sense that they happen within the same nervous system. As a consequence, such interactions should result in measurable behavioral phenomena.² It is likely, however, as Banks (1981) has noted, that both imaginal and perceptual processes are rather more complex than that implied by the demonstrations of Finke and his colleagues. A complete neurophysiological theory of mind needs to encompass both perceptual and cogni-

tive (e.g., imaginal) processes. The task of constructing such a theory is not made easier when the phenomena that must be incorporated in the theory are not capable of being readily and reliably reproduced.

REFERENCES

- BANKS, W. P. (1981). Assessing relations between imagery and perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 844-847.
- BROERSE, J., & CRASSINI, B. (1980). The influence of imagery ability on color aftereffects produced by physically present and imagined induction stimuli. *Perception & Psychophysics*, 28, 560-568.
- BROERSE, J., & CRASSINI, B. (1981). Misinterpretations of imagery-induced McCollough effects: A reply to Finke. *Perception & Psychophysics*, 30, 96-98.
- DE VALOIS, R. L. (1982). Early visual processing: Feature detection or spatial filtering? In D. G. Albrecht (Ed.), *Lecture notes in biomathematics (Vol. 44) Recognition of pattern and form*. New York: Springer-Verlag.
- CRONBACH, L. J. (1970). *Essentials of psychological testing*. New York: Harper and Row.
- FINKE, R. A. (1980). Levels of equivalence in imagery and perception. *Psychological Review*, 87, 113-132.
- FINKE, R. A. (1981). Interpretation of imagery-induced McCollough effects. *Perception & Psychophysics*, 30, 94-95.
- FINKE, R. A., & SCHMIDT, M. J. (1977). Orientation-specific color aftereffects following imagination. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 599-606.
- FINKE, R. A., & SCHMIDT, M. J. (1978). The quantitative measure of pattern representation in images using orientation-specific color aftereffects. *Perception & Psychophysics*, 23, 515-520.
- GINSBURG, A. P. (1982). On a filter approach to understanding the perception of visual form. In D. G. Albrecht (Ed.), *Lecture notes in biomathematics (Vol. 44) Recognition of pattern and form*. New York: Springer-Verlag.
- INTONS-PETERSON, M. J., & WHITE, A. R. (1981). Experimenter naivete and imaginal judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 833-843.
- KAUFMAN, J. H., MAY, J. G., & KUNEN, S. (1981). Interocular transfer of orientation-contingent color aftereffects with external and internal adaptation. *Perception & Psychophysics*, 30, 547-551.
- KOSSLYN, S. M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review*, 88, 46-66.
- KOSSLYN, S. M., PINKER, S., SMITH, G. E., & SHWARTZ, S. P. (1979). On the demystification of mental imagery. *The Behavioral and Brain Sciences*, 2, 535-581.
- KUNEN, S., & MAY, J. G. (1980). Spatial frequency content of visual imagery. *Perception & Psychophysics*, 28, 555-559.
- KUNEN, S., & MAY, J. G. (1981). Imagery-induced McCollough effects: Real or imagined? *Perception & Psychophysics*, 30, 99-100.
- LIEBERMAN, D. A. (1979). Behaviorism and the mind: A (limited) call for a return to introspection. *American Psychologist*, 34, 319-333.
- MAY, J. G., AGAMY, G., & MATTESON, H. H. (1978). The range of spatial frequency contingent color aftereffects. *Vision Research*, 18, 917-921.
- MCCOLLOUGH, C. (1965). Color adaptation of edge detectors in the human visual system. *Science*, 149, 1115-1116.
- MITCHELL, D. B., & RICHMAN, C. L. (1980). Confirmed reservations: Mental travel. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 58-66.

- NATSOULAS, T. (1970). Concerning introspective "knowledge." *Psychological Bulletin*, 73, 89-111.
- NATSOULAS, T. (1978). Consciousness. *American Psychologist*, 33, 906-914.
- RICHMAN, C. L., MITCHELL, D. B., & REZNICK, J. S. (1979a). Mental travel: Some reservations. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 13-18.
- RICHMAN, C. L., MITCHELL, D. B., & REZNICK, J. S. (1979b). The demands of mental travel: Demand characteristics of mental imagery experiments. *The Behavioral and Brain Sciences*, 2, 564-565.
- STANLEY, G., & HOFFMAN, W. C. (1976). Orientation-specific color effects without adaptation. *Bulletin of the Psychonomic Society*, 7, 513-514.

NOTES

1. After finishing the tasks outlined, all observers were requested to complete three short self-report questionnaires of imagery ability. These data were not directly relevant to the present investigation and will appear in a subsequent report.

2. Our discussion is based on attempts to generate IICAEs using grating stimuli. Kaufman, May, and Kunen (1981) and Kunen and May (1980) have reported IICAEs using checkerboard stimuli which we have deliberately not discussed in this paper. Our reasons for not discussing these reports are:

(a) The use of checkerboard stimuli relates to the debate between those who argue for a feature-detector approach to understanding visual perception and those who argue for an approach based on Fourier analysis (e.g., see De Valois, 1982; Ginsburg, 1982). The debate remains unresolved and is fueled by complex and difficult-to-reconcile data from checkerboard studies. Compare, for example, the results reported by May, Agamy, and Matteson (1978) in Experiments 1 and 2 using spatial frequencies of less than 3.0 cycles/deg. If the exact nature of processing of physically present checkerboard stimuli is unknown (or at best not the subject of general agreement), then it seems premature to us to use such stimuli in studies attempting to clarify the relationship between perception and imagination.

(b) Checkerboard IICAEs are the *opposite* of checkerboard CAEs (at least for checkerboards having a fundamental spatial frequency of 3.0 cycles/deg; Kunen & May, 1980). Kunen and May explain this difference by saying that when observers are asked to imagine a checkerboard, they perform a high-pass filtering operation such that only high spatial frequencies remain in the image (and therefore only high spatial frequency mechanisms are adapted). This "filtering" is unconscious, and paradoxical in that (as far as we can ascertain) the observers don't report that their imagined checkerboard is different in "appearance" from the real checkerboard used as a prompt. We have already argued that one of the problems in equating IICAEs and CAEs in terms of underlying mechanisms is that there are fundamental and important differences in their properties (Broerse & Crassini, 1980, 1981). The

differences between checkerboard IICAEs and checkerboard CAEs described above lend further support to this argument.

For these reasons, we feel that checkerboard IICAEs raise a special set of issues distinct from those that have to do with the grating IICAEs discussed in the present paper.

APPENDIX

Instructions for Color Reports

Observers in FC conditions were asked to report which half of the test slide appeared more red, and to record this in the appropriate part of the response blank (i.e., top or bottom; left or right), corresponding to the test stimulus in question. If this discrimination was too difficult (e.g., only green was seen), the observers were asked to choose which half appeared less green. If they could not make any color discriminations at all, then they were to guess. Magnitude-estimation observers were briefly (2-3 sec) shown homogeneous patches of the red and green colors used during inspection and told that these colors were to be the standards (with a score of 100) for all subsequent color judgments. Judgments were to be made about each half of the test slide separately, and involved first determining whether there was any color appearance (no color present was to be given a score of 0). If the test slide half appeared colored, the observers were to classify the color as red or green and estimate the "strength" of the color on a scale of 0 to 100.

Imagery Instructions

The instructions to imagine contours were: "When the red color is presented, I would like you to visualise as vividly as you can that same set of vertical bars that you just observed. Make these bars as vivid and as real as you can, having them completely embedded in the red so that all you can see is a pattern of black and red vertical bars. Likewise, when the green color is presented, I would like you to visualise as vividly as you can that same set of horizontal bars that you just observed. Make these bars as vivid and real as you can, having them completely embedded in the green so that all you can see is a pattern of black and green horizontal bars. Try to the very best of your ability to imagine vividly the presence of these suggested bar orientations. To help you imagine the appropriate bar orientations each time, I shall say the word 'vertical' or 'horizontal' before each slide is presented, to indicate the bar orientation to visualize on the slide presented thereafter."

(Manuscript received May 11, 1983;
revision accepted for publication December 5, 1983.)